Aus der Klinik und Poliklinik für Nuklearmedizin der
Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. P. Bartenstein

Auswirkung von Referenzregionen, Partialvolumeneffekten und Segmentierungsartefakten auf die Quantifizierung der Positronen-Emissions-Tomographie bei der Alzheimer-Erkrankung

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Marcus Högenauer
aus
Landshut

Jahr
2017

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. med. Axel Rominger
Mitberichterstatter:	Prof. Dr. med. Oliver Pogarell Priv. Doz. Dr. med. Johannes Levin
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	05.10.2017

Die vorliegende kumulative Dissertation umfasst zwei bereits publizierte Manuskripte:

Högenauer, M.*, Brendel, M.*, Delker, A., Därr, S., Weiss, M., Bartenstein, P., Rominger, A., for the Alzheimer's Disease Neuroimaging Initiative (2016). **Impact of MRI-based Segmentation Artifacts on Amyloid- and FDG-PET Quantitation.** Curr Alzheimer Res. 2016 13(5): 597-607.

* Geteilte Erstautorenschaft

Brendel, M., **Högenauer, M.**, Delker, A., Sauerbeck, J., Bartenstein, P., Seibyl, J., Rominger, A., for the Alzheimer's Disease Neuroimaging Initiative (2015). **Improved longitudinal [^{18}F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction.** Neuroimage. 2015 108: 450-459.

Kumulative Dissertation gemäß § 4a der Promotionsordnung

Inhalt

1. Einführung

Aktuell leben in Deutschland etwa 1,6 Millionen Menschen mit einer demenziellen Erkrankung. Mit einem Anteil von mindestens zwei Drittel der Betroffenen ist die Alzheimer-Erkrankung die häufigste Demenzform (Bickel 2016). Dem World Alzheimer Report 2015 zufolge gab es 2015 46,8 Millionen Demenzerkrankte weltweit, infolge der steigenden Lebenserwartung gehen die Prognosen von bis zu 131,5 Millionen Demenzpatienten im Jahr 2050 aus. Neben der stetig zunehmenden Anzahl an Einzelschicksalen, die sich mit dem Verlust der Lebensqualität und der persönlichen Integrität arrangieren müssen, gilt es, eine enorme gesamtgesellschaftliche Belastung infolge drastisch steigender Gesundheits- und Sozialausgaben zu bewältigen (Alzheimer's Disease International 2015). Jedoch könnten bereits kleinere therapeutische Erfolge, die zu einem verzögerten Beginn und verminderter Progression von ein bis zwei Jahren führen, zu einer deutlichen Reduktion der global prognostizierten Alzheimer-Prävalenz und der daraus resultierenden Belastung beitragen (Brookmeyer, Johnson et al. 2007). Die aktuell zur Verfügung stehenden Therapiemöglichkeiten weisen nur bescheidene Erfolge auf und beschränken sich zudem auf eine symptomatische Behandlung (Hugo and Ganguli 2014, Gauthier, Albert et al. 2016). Molekularpathologische Abläufe zu modulieren und den Beginn und die Progression neurodegenerativer Erkrankungen zu verhindern ist das Ziel künftiger Therapeutika. Eine robuste neurologische Bildgebung wird somit in zweierlei Hinsicht immer wichtiger: Zunächst zur Differenzierung zwischen asymptomatischen Stadien und Vorläuferstufen neurodegenerativer Erkrankungen und im Folgenden zum Monitoring

neuentwickelter molekularmodulierender Therapeutika (Hampel, Prvulovic et al. 2011).

1.1 Alzheimer`s Disease Neuroimaging Initiative (ADNI)

ADNI ist ein durch öffentliche und private Einrichtungen finanziertes Forschungsprojekt, das Anfang des Jahrtausends initiiert wurde, um die Forschungsbemühungen im Rahmen der Alzheimer-Erkrankung zu bündeln und voranzutreiben. Ziel der Initiative war zunächst die Entwicklung optimierter Methoden, um einheitliche Standards für die bildgebende Datenerhebung in longitudinalen und multizentrischen Studien sicherzustellen. Die erarbeiteten Methoden sollten bei der Erhebung longitudinaler, funktioneller und struktureller Bilddaten einer großen Kohorte an älteren Kontrollpersonen und Probanden mit leichter kognitiver Einschränkung und Alzheimer-Demenz Anwendung finden und erfolgte an klinischen Standorten in den USA und Kanada. Mit Hilfe der neben den Bilddaten zugleich ermittelten klinischen, kognitiven, genetischen und biochemischen Biomarkern sollten charakteristische Zusammenhänge der Erkrankung vom Prodromalstadium bis hin zum voll ausgeprägten Krankheitsbild ermittelt und leistungsstarke Messgrößen zur Diagnostik und Messung von Therapieeffekten identifiziert werden. Um der Alzheimer-Forschung Vorschub zu leisten, war es den Initiatoren ein Anliegen, die generierte Datenbank Wissenschaftlern weltweit zugänglich zu machen (Mueller, Weiner et al. 2005, Mueller, Weiner et al. 2005, Weiner, Aisen et al. 2010).

1.2 Cerebrales β-Amyloid

Cerebrale Amyloidablagerungen, sogenannte extrazelluläre ß-Amyloid (Aß)-Plaques, sind neben Tau-Ablagerungen charakteristische Zeichen einer Alzheimer-Erkrankung (Duyckaerts, Delatour et al. 2009). Aß-Peptid ist ein natürlich vorkommendes Spaltprodukt der ß- und γ-Sekretase aus dem Amyloid-Vorläufer-Protein (APP) und variiert in seiner Länge zwischen 36 und 43 Aminosäuren. Das vorherrschende Peptid ist dabei $Aß_{40}$. Im physiologischen Gleichgewicht wirkt Aß modulierend auf synaptische Exzitationen ein. Ist das Gleichgewicht zwischen Entstehung und Abbau von Aß jedoch gestört, kommt es laut Amyloidhypothese zur Bildung von Oligomeren und Fibrillen und schließlich zur Aggregation der typischen Plaques. Insbesondere das Peptid $Aß_{42}$ scheint diese Plaquebildung zu begünstigen und zugleich schädigenden Einfluss auf Nervenzellen zu nehmen. Neben den extrazellulären Ablagerungen lassen sich auch intrazellulär Aggregate aus Tau-Proteinen nachweisen. Tau ist ein lösliches Protein, das zahlreich intrazellulär vorhanden ist und wesentlich zur Bildung und Stabilisierung von Mikrotubuli beiträgt. Durch Hyperphosphorylierung geht die Affinität zu den Mikrotubuli verloren, was zur Instabilität derselben führt und den axonalen Transport in den Nervenzellen einschränkt. Bei Alzheimer-Patienten (AD) konnte bisher keine genetische Komponente der Tau-Mutation festgestellt werden (Querfurth and LaFerla 2010). In anderen neurodegenerativen Erkrankungen wie z.B. der frontotemporalen Demenz mit Parkinsonismus ist dies jedoch der Fall (Goedert and Jakes 2005). Vielmehr deutet alles daraufhin, dass Aß-Anreicherungen unter anderem zu vermehrter Bildung von Tau-Neurofibrillen führen können (Gotz, Chen et al. 2001). Die Hypothese der Amyloidkaskade gründet auf konkreten Mutationen

der Gene APP und Presenilin 1 und 2, die an der Aß-Synthese beteiligt sind und in direkter Assoziation zu der früh beginnenden familiären Alzheimer-Demenz stehen. Die verantwortlichen Faktoren, die zur Entstehung der Aß-Plaques in der wesentlich häufigeren sporadischen Alzheimer-Demenz, mit Beginn nach dem 65. Lebensjahr, führen, sind derzeit noch nicht abschließend geklärt (Winblad, Amouyel et al. 2016). Ein deutlich erhöhtes Lebenszeitrisiko zur Entwicklung einer sporadischen Alzheimer-Demenz konnte bisher für Träger eines ε4 Alleles des für den Cholesterintransport verantwortlichen Apolipoprotein E Gens nachgewiesen werden (Poirier, Davignon et al. 1993). Diese zeigten ebenfalls vermehrte Amyloidablagerungen wegen vermutlich zu geringem Aß-Abbau. Zudem deuteten Forschungen eine erhöhte neuronale Toxizität von Aß im Zusammenhang mit Apolipoprotein E ε4 an (Winblad, Amouyel et al. 2016). Eine Trägerschaft des Apolipoprotein E ε2 Alleles konnte hingegen als protektiver Faktor zur Entwicklung einer Alzheimer-Demenz identifiziert werden (Corder, Saunders et al. 1994). Um die Plaqueentstehung einzudämmen, gibt es verschiedenste Lösungsansätze in Form von γ-Sekretase Inhibitoren, Impfungen mit Aß, monoklonalen Antikörpern gegen diverse Aß-Epitope und Aggregationshemmern, deren Wirksamkeit Gegenstand aktueller Forschung in klinischen Studien ist. Dabei treten jedoch teilweise erhebliche Nebenwirkungen auf (Querfurth and LaFerla 2010). Dies zeigt, dass bei der Suche nach Zielstrukturen für Behandlungsansätze, wie beispielsweise der Enzyme ß- und γ-Sekretase, die neben den amyloidogenen Stoffen weitere Substrate metabolisieren, eine selektive Hemmung der Aß-Bildung erfolgen muss, um eine ansonst korrekte biologische Funktion aufrechtzuerhalten. Aus diesem Grund ist die

Erforschung der multifaktoriellen Genese der Aß- und Tau-Plaquebildung und deren Auswirkung auf die klinische Ausprägung der Alzheimer-Demenz weiterhin von größter Bedeutung, um zielgerichtete Behandlungstherapien zu ermöglichen. Gleichzeitig gilt es aber auch geeignete Biomarker zu entwickeln und zu standardisieren, um den Therapieverlauf künftiger präventiver und krankheitsmodulierender Arzneimittel zu überwachen (Winblad, Amouyel et al. 2016).

1.3 Positronen-Emissions-Tomographie bei Alzheimer-Erkrankung

Die Positronen-Emissions-Tomographie (PET) ermöglicht es, molekulare Veränderungen im Rahmen der Alzheimer-Erkrankung *in vivo* darzustellen. Durch Verwendung verschiedener radioaktiv markierter Trägerstoffe (Radiotracer) kann zum einen die Amyloidpathologie und zum anderen die degenerativ veränderte neuronale Aktivität abgebildet werden. Die daraus gewonnenen Erkenntnisse können wertvolle Ergänzungen zu den klinischen Untersuchungen liefern und unterstützend zur Diagnosefindung und Prognose von neurodegenerativen Demenzen beitragen (Nasrallah and Wolk 2014). In einer longitudinalen Studie stellten Jack et al, die lange vor dem Auftreten kognitiver Defizite, primär im Alter zwischen 60 und 75 Jahren, dynamisch ansteigende Amyloidanreicherungen fest, deren Dynamik um das 75. Lebensjahr abnimmt und nahezu ein Plateau erreicht (Jack, Therneau et al. 2016). In einem postulierten Modell dynamischer Biomarker tritt die neuronale Degeneration deutlich später in Erscheinung und korreliert dann jedoch stark mit klinischen Symptomen (Jack, Knopman et al. 2010). Villemagne et al sind mit dieser Erkenntnis und These konform. Sie beobachteten mit Hilfe des [^{11}C]-Pittsburgh compound B ([^{11}C]-PiB) einen

über zwei Jahrzehnte andauernden und im späten Stadium verringerten dynamischen Prozess der Amyloidanreicherung. Die Möglichkeit zur Detektion dieser Pathologie viele Jahre vor dem Auftreten typischer Veränderungen in der Kognition und am Hirnvolumen eröffnet ein großes Zeitfenster zur Anwendung antiamyloider Therapien (Villemagne, Burnham et al. 2013). Auf Grund der geringen Halbwertszeit von [^{11}C] markierten Radiotracern von ca. 20 Minuten ist die Einsetzbarkeit des wohl am meist untersuchten Amyloidtracers [^{11}C]-PiB auf Forschungseinrichtungen mit eigenem Zyklotron beschränkt. Eine verbesserte klinische Anwendbarkeit erlauben hingegen [^{18}F] markierte Tracer, wie z.B. [^{18}F]-Florbetapir ([^{18}F]-AV45), mit einer deutlich verlängerten Halbwertszeit von ca. 110 Minuten. Beide Radiotracer führen insgesamt zu analogen Ergebnissen (Wolk, Zhang et al. 2012). Ein Amyloidnachweis mittels [^{18}F]-AV45 PET gegenüber einer Autopsie ergab ebenfalls signifikante Übereinstimmungen in den Resultaten und validierte somit die [^{18}F]-AV45 Bildgebung als nützliches Instrument zur Bestimmung von Amyloidplaques *in vivo* (Clark, Pontecorvo et al. 2012). Unter Verwendung des stabilen Biomarkers [^{18}F]-Fluorodesoxyglucose ([^{18}F]-FDG) lässt sich die synaptische Aktivität des Gehirns darstellen (Segobin, La Joie et al. 2015). Noch vor dem Auftreten kognitiver Auffälligkeiten ermöglicht [^{18}F]-FDG die Detektion eines verringerten Metabolismus in Folge neurodegenerativer Veränderungen (de Leon, Convit et al. 2001, Jagust, Gitcho et al. 2006). Ein reduzierter Stoffwechsel *per se* ist kein spezifischer Hinweis auf das Vorliegen einer Alzheimer-Demenz. Im Verlauf der Erkrankung sind jedoch charakteristische Regionen des Gehirns wie Precuneus, posteriorer cingulärer Cortex, Hippocampus, Anteile des Parietal- sowie Temporallappens und in

fortgeschrittenen Fällen der präfrontale Cortex von schrittweisen Verschlechterungen der metabolischen Defizite betroffen (Johnson, Fox et al. 2012). Während bei bestehender Alzheimer-Demenz die Amyloidbelastung auf hohem Niveau relativ stabil bleibt, korreliert bei Progression der Erkrankung der abnehmende Glucosestoffwechsel signifikant mit den progredienten kognitiven Defiziten (Engler, Forsberg et al. 2006). Die beschriebenen bildgebenden Biomarker haben nicht nur das Potenzial, die Diagnostik zu präzisieren und den Krankheitsverlauf zu überwachen, sondern scheinen auch als Outcome-Parameter in therapeutischen Studien geeignet zu sein. Im Gegensatz zu gängigen kognitiven und klinischen Parametern erlauben sie kleinere Studienpopulationen und kürzere Beobachtungszeiträume (Mueller, Weiner et al. 2005). Dies wird zu steigenden Anforderungen an die molekulare Bildgebung führen, um beispielsweise Veränderungen am Gehirn, die mit klinischen Symptomen korrelieren, nach therapeutischen Interventionen zu bestimmen (Nordberg 2011). Aus diesem Grund gilt es, die Genauigkeit der Bildgebung weiter zu verbessern, um auch minimalste Veränderungen zuverlässig detektieren zu können.

1.4 Referenzregion

Die ermittelten Standard-Uptake-Values (SUV) einer PET werden gewöhnlich mit Hilfe einer Referenzregion normalisiert. Dies ermöglicht eine verbesserte Vergleichbarkeit der erfassten Werte eines Probanden im Rahmen wiederholter Untersuchungen oder zwischen verschiedenen Probanden (Yakushev, Landvogt et al. 2008). Zudem können die Varianzen der ermittelten Messwerte verringert werden (Minoshima, Frey et al. 1995).

Um adäquate SUV-Ratios (SUVR) zu erhalten, kommt der Referenzauswahl eine wesentliche Bedeutung zu. Laut Yakushev et al sollte sie zuverlässig ermittelt werden können, bestenfalls nicht oder nur geringfügig von degenerativen Prozessen betroffen sein, sollte nach Möglichkeit nicht durch äußere physiologische Stimuli beeinflusst werden und die Traceraufnahme sollte bei gesunden Kontrollpersonen (HC) keine signifikanten Abweichungen aufweisen (Yakushev, Landvogt et al. 2008). Eine ideale Referenz sollte zudem eine präzise Differenzierung zwischen den Diagnosegruppen ermöglichen (Dukart, Mueller et al. 2010). Während für [^{18}F]-FDG Untersuchungen verschiedenste Referenzregionen wie Cerebellum (CBL), Pons/Hirnstamm, Global Mean und Reference Cluster einen weitreichenden Diskurs erfahren haben (Yakushev, Landvogt et al. 2008, Borghammer, Cumming et al. 2009, Bohnen, Djang et al. 2012, Dukart, Perneczky et al. 2013), wird für die [^{18}F]-AV45 und [^{11}C]-PiB Amyloidanalysen weitgehend nur die graue Hirnsubstanz (GM) des CBL oder das komplette CBL als Referenz verwendet (Weiner, Veitch et al. 2013). Eine longitudinale [^{11}C]-PiB Studie von van Berckel et al mit leicht kognitiv eingeschränkten und an Alzheimer erkrankten Probanden zeigte für die mit der GM des CBL normalisierten Werte eine starke interindividuelle Variabilität (van Berckel, Ossenkoppele et al. 2013). Dies lässt darauf schließen, dass die oben definierten Anforderungen an eine ideale Referenzregion im CBL noch nicht gänzlich erfüllt sind und es weitere Verbesserungen für die Quantifizierung von Amyloid mittels PET zu erforschen gilt.

1.5 Partialvolumeneffekt/-korrektur

Die räumliche Auflösung der Positronen-Emissions-Tomographen ist auf Grund technischer Begebenheiten begrenzt (Erlandsson, Buvat et al. 2012). Bei Strukturen mit ähnlichen oder kleineren Abmessungen, verglichen mit der Auflösung des Positronen-Emissions-Tomographen, treten geringere Zählraten (Counts) auf. Der Schwund offensichtlich vorhandener Aktivität und die damit verbundene Abschwächung der erwarteten Signalintensität wird als Partialvolumeneffekt (PVE) bezeichnet (Hutton and Osiecki 1998). Es gibt zwei Aspekte des PVE, zum einen eine Streuung der Radioaktivität einer Zielstruktur in die Umgebung bzw. von der Umgebung in die Zielstruktur auf Grund der Point-Spread Function (PSF) des Systems. Diese ist ein Maß für die räumliche Auflösung des Systems und wird mit Hilfe des Full-Width-at-Half-Maximum (FWHM) genauer bestimmt (Erlandsson, Buvat et al. 2012). Ist die zu untersuchende Struktur kleiner als das Zweifache des FWHM wird der PVE ein besonders beeinflussender Faktor (Segobin, La Joie et al. 2015). Als zweiter Aspekt führt das relativ große Voxelvolumen dazu, dass ein Voxel, bevorzugt an den Rändern der Zielstruktur, zwei oder mehr Gewebearten (GM, weiße Hirnsubstanz (WM) oder Liquor (CSF)) enthält. Dieses Phänomen wird als „tissue-fraction effect“ bezeichnet. Auf Grund des relativ dünnen, nur wenige Millimeter dicken, cerebralen Cortex ist in der neuronalen PET-Bildgebung mit ausgeprägten PVE zu rechnen (Erlandsson, Buvat et al. 2012), die durch neurodegenerative Atrophie, wie sie bei AD auftritt (Thompson, Hayashi et al. 2003), verstärkt werden. Etablierte Korrekturmethoden, wie Geometric-Transfer-Matrix (GTM) (Rousset, Ma et al. 1998) oder Müller-Gärtner-Methode (Muller-Gartner, Links et al. 1992), verwenden zur Rekonstruktion der radioaktiven Verteilung

zusätzliche strukturelle Informationen aus Magnet-Resonanz-Tomographie (MRT)-Daten in Form von Segmentierungen und profitieren von verbesserten Ergebnissen (Erlandsson, Buvat et al. 2012). Trotz intensivster Bemühungen, die automatisierten Segmentierungen zu verbessern (Smith 2002, Segonne, Dale et al. 2004, Mikheev, Nevsky et al. 2008, Leung, Barnes et al. 2011), sind gänzlich fehlerfreie Segmentierungen derzeit nicht verfügbar. Die Auswirkungen fehlerhafter Segmentierungen auf die Quantifizierung von PET-Daten unter Verwendung der Partialvolumeneffektkorrektur (PVEC) sind noch nicht in vollem Umfang erforscht.

2. Inhalte der Promotionsarbeit

2.1 Referenzregionen und Partialvolumeneffektkorrektur – Einfluss auf longitudinale Amyloid PET-Untersuchungen

Brendel, M., **Högenauer, M.**, Delker, A., Sauerbeck, J., Bartenstein, P., Seibyl, J., Rominger, A., for the Alzheimer's Disease Neuroimaging Initiative (2015). **Improved longitudinal [^{18}F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction.** Neuroimage. 2015 108: 450-459.

Im Rahmen meiner Zweitautorenschaft „Improved longitudinal [^{18}F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction“ (Brendel, Hogenauer et al. 2015) wurde die Auswirkung der Auswahl unterschiedlicher Referenzregionen auf die Quantifizierung von [^{18}F]-AV45 PET-Ergebnissen untersucht, sowie der Einfluss einer Volume-Of-Interest (VOI)-basierten PVEC analysiert. Die hierfür verwendeten Datensätze, bestehend aus Amyloid PET und T1-gewichteten MRT-Bildern, wurden aus der ADNI-Datenbank bezogen und mit dem Programm PMOD PNEURO V. 3.5 weiter prozessiert. Dazu erfolgte im ersten Schritt eine rigide Co-Registrierung der PET-Datensätze mit ihren jeweiligen MRT-Bildern, um eine lineare Transformation (PET-2-MRT) zu erhalten. Desweiteren generierten wir eine nichtlineare Transformation (MRT-2-MNI), indem die individuellen MRT-Daten auf den standardisierten Montreal Neurological Institute (MNI)-Atlas co-registriert wurden. Die Kombination beider Transformationen ermöglichte es schließlich, die individuellen PET-Bilddaten in den standardisierten MNI-Raum zu überführen. Eine Segmentierung der MRT-Datensätze in GM und WM sowie CSF (Ashburner and Friston 2005) erlaubte es, in Anlehnung an den Hammers Atlas (Hammers, Allom et al. 2003), für jeden Probanden 83 individuelle VOIs im

MNI-Raum zu generieren. Im Anschluss wurde die Korrektheit der automatisierten Co-Registrierungen und Segmentierungen (Baseline N=1018/Follow-up N=278) visuell überprüft. Fehlerhafte Co-Registrierungen (Baseline N=36 (3,5%)/Follow-up N=8 (3%)) konnten durch eine halbautomatische Bearbeitung wieder der Auswertung zugeführt werden. Fehlerhafte und durch Artefakte behaftete Segmentierungen (Baseline N=56 (5,5%)/Follow-up N=20 (7,2%)) mussten jedoch ausgeschlossen werden, da eine Veränderung der Standardparameter zu einer Verfälschung der VOI-Ergebnisse hätte führen können.

Die folgenden Auswertungen der GM-Volumina wurden im MNI-Raum durchgeführt, um den verschiedenen Hirnvolumina gerecht zu werden. Im Gegensatz dazu erfolgten die VOI-Analysen im individuellen PET-Raum, indem die Hammers VOIs durch eine Umkehrung der (PET-2-MNI) Transformation in den nativen PET-Raum konvertiert wurden. Dieser fand in der nach der GTM-Methode durchgeführten VOI-basierten PVEC ebenfalls Anwendung (Rousset, Ma et al. 1998, Rousset, Collins et al. 2008). Der angewendete Korrekturalgorithmus ließ dabei für die Berechnung sowohl den Background als auch die WM-Regionen einfließen, um speziell für die Fragestellung der Amyloid Quantifizierung adäquate Ergebnisse zu erzielen (Thomas, Erlandsson et al. 2011, Brendel, Delker et al. 2014).

Für die Analysen wurde zunächst eine kombinierte (COMP)-VOI, bestehend aus Frontal-, Parietal- und Temporalcortex sowie dem precunealen/posterioren cingulären Cortex, definiert. Die COMP-VOIs wurden weiter differenziert in eine vollständige Atlas-Region (COMP-FULL-VOI) und in segmentierte GM-Regionen ohne (COMP-GM-VOI) und mit Partialvolumeneffektkorrektur (COMP-PVEC-GM-VOI). Jede dieser VOIs

wurde mit den drei Referenzgeweben CBL, Hirnstamm (BST) und WM normalisiert. Die detaillierten Arbeitsschritte zeigt Abbildung 1. Die neun entstandenen Konfigurationen wurden durch eine Kontroll-SUVR, bestehend aus COMP-GM-VOI normalisiert mit cerebellärer FULL-VOI, ergänzt (Joshi, Pontecorvo et al. 2012, Johnson, Sperling et al. 2013).

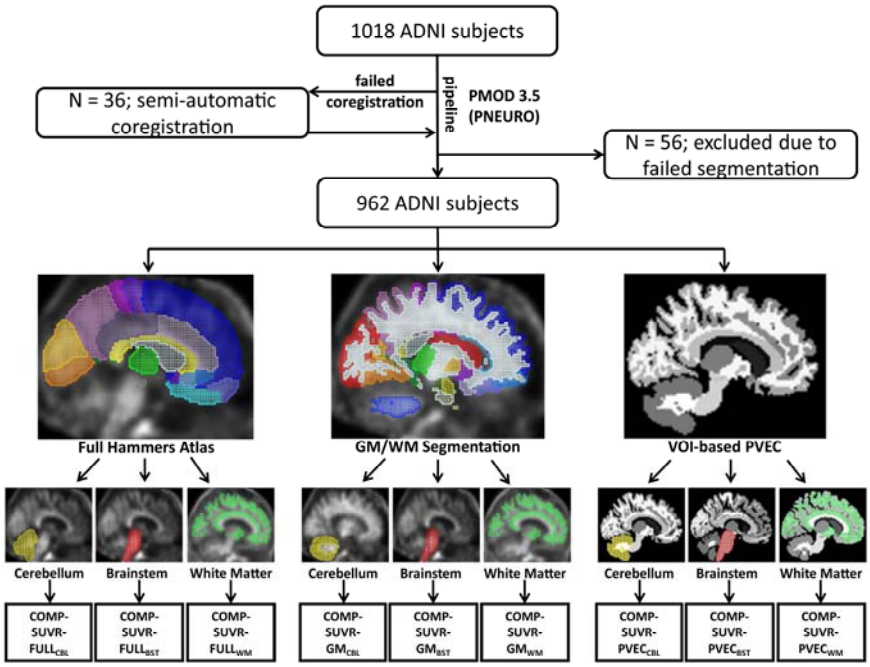


Abbildung 1: Arbeitsschritte von der ADNI-Datenbank zur VOI-basierten Analyse. Alle Probanden, die zur Baseline eine [^{18}F]-AV45 PET und eine T1-gewichtete MRT erhalten hatten, wurden vollautomatisch mit PNEURO (PMOD 3.5) aufbereitet. Im Anschluss erfolgte eine visuelle Kontrolle der Ergebnisse, bei fehlerhaften Co-Registrierungen erfolgte eine manuelle Nachjustierung. Segmentierungsfehler und ausgedehnte Artefakte wurden gänzlich ausgeschlossen. Die verbleibenden Probanden wurden unter Verwendung von kompletten Atlas VOIs (Hammers, Allom et al. 2003), segmentierten GM VOIs und segmentierten GM VOIs nach PVEC weiter analysiert. Die resultierenden VOI-Messwerte wurden entweder mit Hilfe des CBL (gelb), des BST (rot) oder der WM (grün) normalisiert.

In der Referenzregion WM stellten wir diagnoseübergreifend die stabilsten SUVs ohne signifikante Unterschiede fest, wohingegen sich beim kompletten CBL (CBL-FULL) und CBL-GM, am stärksten aber bei BST, signifikante SUV-Abnahmen zwischen HC und AD von bis zu 10,4% ergaben ($p<0,001$).

Die nach Diagnosegruppen unterteilten COMP-SUVR-Durchschnittswerte der zehn Konfigurationen ergaben im Vergleich hoch signifikante Unterschiede ($p<0,001$) zwischen leicht kognitiv Beeinträchtigten (MCI)/HC sowie AD/HC, unabhängig von Segmentierung, PVEC und der Wahl des Referenzgewebes. Jedoch zeigte sich eine deutliche Erhöhung der prozentualen Unterschiede zwischen den Diagnosegruppen durch gleichzeitige Anwendung von Segmentierung und PVEC, insbesondere in Kombination mit BST und WM als Referenzregion auf bis zu 22% für MCI/HC und 70% für AD/HC. Allgemein führten die Referenzregionen BST und WM zu niedrigeren SUVRs im Vergleich zur Normalisierung mit CBL. Mit Hilfe einer Receiver-Operating-Characteristic (ROC)-Analyse konnte zudem eine deutlich verbesserte Trennschärfe zwischen den Diagnosegruppen AD/HC und MCI/HC bei Verwendung von BST und WM als Referenzregionen in Verbindung mit PVEC gezeigt werden.

Das Ausmaß der PVE wurde durch die Korrektur der COMP-GM-VOI-Konfiguration ohne Normalisierung durch ein Referenzgewebe eruiert. Die prozentualen Abweichungen zwischen den Diagnosegruppen unterschieden sich dabei in der COMP-Konfiguration hoch signifikant ($F=166$; $p<0,001$). Dies zeigte sich in mittleren SUV-Reduzierungen bei HC von -18% und MCI von -10%, bei AD jedoch in einem durchschnittlichen SUV-Anstieg von 7%, was im Einklang mit der signifikanten Atrophie der AD (mittleres GM Volumen: $452cm^3$) ($p<0,001$) im Gegensatz zu HC/MCI (mittlere GM Volumina: $485cm^3$/$480cm^3$) stand. Die PVEC bei CBL-GM ergab diagnoseübergreifend reduzierte SUVs (HC: -8%; MCI: -7%; AD: -6%).

Im longitudinalen Verlauf gab es nur einen marginalen Anstieg der durchschnittlichen Referenzregion SUVs in der Größenordnung von

0,1-0,3%. In den COMP-Konfigurationen zeigten sich unter Verwendung der PVEC im zeitlichen Verlauf die größten Anstiege im mittleren [^{18}F]-AV45 Uptake von bis zu 4,6%, und das unabhängig der ausgewählten Referenzregion. Die Variabilität zwischen den Probanden war jedoch im longitudinalen Verlauf für WM am geringsten, gefolgt von BST und am stärksten bei CBL.

Zusammenfassend ließ sich feststellen, dass die Verwendung von VOI-basierter PVEC in Verbindung mit WM oder BST als Referenzregionen die Trennschärfe zwischen HC und MCI deutlich verbessert hat, welche in künftigen Behandlungsansätzen die Zielgruppen darstellen werden. Ebenso zeigte sich durch die PVEC und WM eine deutlich genauere Erfassung der longitudinalen Amyloidogenese.

2.2 Artefakte bei MRT-basierten Segmentierungen und deren Einfluss auf Amyloid und FDG PET-Quantifizierungen

Högenauer, M., Brendel, M., Delker, A., Därr, S., Weiss, M., Bartenstein, P., Rominger, A., for the Alzheimer's Disease Neuroimaging Initiative (2016). **Impact of MRI-based Segmentation Artifacts on Amyloid- and FDG-PET Quantitation.** Curr Alzheimer Res. 2016 13(5): 597-607.

Im Rahmen der in 2.1 beschriebenen Studie wurden nach der Bilddatenaufbereitung mit Hilfe eines automatisierten Protokolls die Ergebnisse visuell auf korrekte Co-Registrierung und Segmentierung überprüft. Dabei wurden bei 76 Bilddatensätzen komplett fehlgeschlagene beziehungsweise mit deutlichen Artefakten behaftete Segmentierungen identifiziert. Da zu diesem Zeitpunkt die Auswirkungen von Segmentierungsartefakten auf PET-Quantifizierungen noch nicht evaluiert waren und eine Abweichung von den Parametern des

Aufbereitungsprotokolls zu nicht absehbaren Verfälschungen der VOI-Ergebnisse geführt hätte, mussten die betroffenen Bilddatensätze von den weiteren Analysen ausgeschlossen werden.

In der Folgearbeit „Impact of MRI-based Segmentation Artifacts on Amyloid- and FDG-PET Quantitation“ (Hogenauer, Brendel et al. 2016) befassten wir uns nun näher mit dieser Thematik und stellten fest, dass ausgedehnte Artefakte, insbesondere bei Verwendung einer PVEC, demaskiert und beseitigt werden müssen, da sie einen beträchtlichen Einfluss auf PET-Quantifizierungen haben.

In dieser Arbeit wurden im ersten Schritt die von ADNI bezogenen und vorbearbeiteten Daten, bestehend aus [^{18}F]-AV45/[^{18}F]-FDG PET-Bilddatensätzen und T1-gewichteten MRT-Daten, co-registriert und segmentiert. Dies erfolgte analog dem Vorgehen in der obigen Publikation mit PMOD PNEURO V. 3.5. Lineare, rigide Co-Registrierung der beiden PET-Datensätze auf das individuelle MRT-Bild (PET-2-MRT) und nicht lineare Transformation der individuellen MRT-Daten in den MNI-Raum (MRT-2-MNI). Unter Verwendung der beiden Transformationen (PET-2-MRT) und (MRT-2-MNI) wurden beide PET-Datensätze in den MNI-Raum transformiert. Die Segmentierung erfolgte ebenfalls in die Kompartimente GM, WM und CSF (Ashburner and Friston 2005). Im Anschluss wurden 83 VOIs nach Hammers (Hammers, Allom et al. 2003) im MNI-Raum generiert und für jeden dieser VOIs ein zusätzlicher PVE-korrigierter Wert (Rousset, Ma et al. 1998, Rousset, Deep et al. 2000, Rousset, Collins et al. 2008) berechnet. Jeder der 119 Probanden verfügte über eine zweijährige Follow-up Aufnahme. Die Segmentierungen der 238 MRT-Datensätze wurden visuell

auf Artefakte überprüft. 66 Segmentierungen wiesen mindestens ein Artefakt auf und wurden daraufhin mit ihrem individuellen [^{18}F]-FDG PET-Binärbild maskiert, um die Artefakte zu entfernen (Abbildung 2). Die korrigierten Segmentierungen wurden erneut mittels PMOD prozessiert, um nun die korrekten VOI- und Volumenergebnisse zu erhalten.

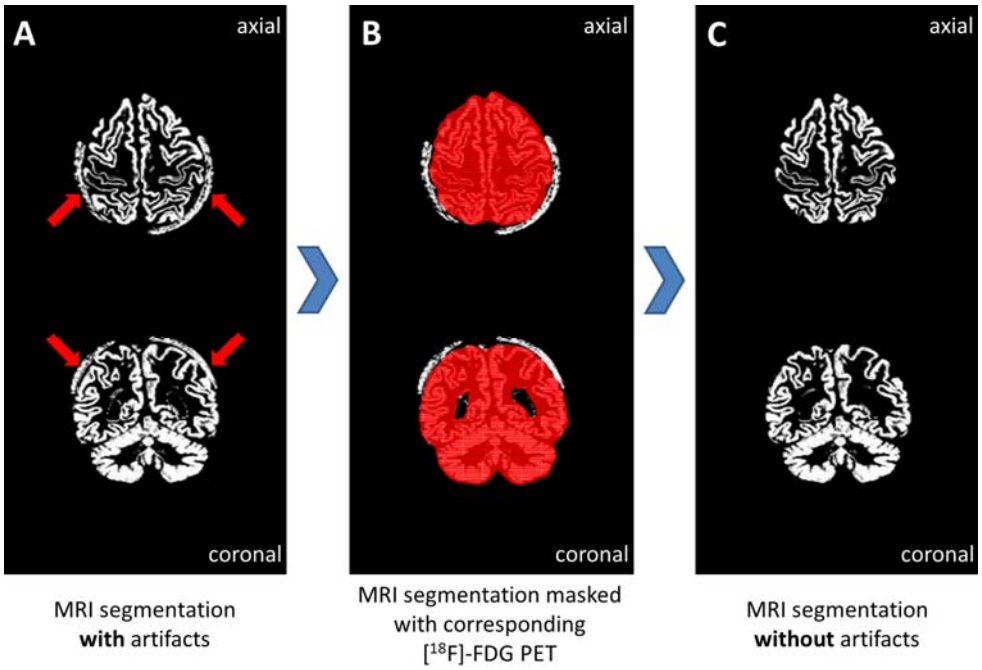


Abbildung 2: Arbeitsablauf der Artefaktmaskierung. **(A)** Beispiel eines Segmentierungsartefakts (hervorgehoben durch rote Pfeile) in axialer und coronarer Ebene. **(B)** Eine bestmögliche Anpassung des Binärbildes des zugehörigen [^{18}F]-FDG PET an die Gehirnränder wurde mit Hilfe eines individuellen Schwellenwertes (20-37% der absoluten Aktivität) erreicht. **(C)** Artefaktfreie MRT-Segmentierung nach Entfernung von Voxeln außerhalb der Maske.

Für die weiteren Analysen wurden sechs GM-VOIs definiert: Frontal-, Parietal-, Temporal- und posteriorer cingulärer Cortex, sowie eine Composite-VOI, bestehend aus den vier vorgenannten Cortexarealen, und CBL. Die größten prozentualen SUV-Fehler ergaben sich Tracer-unabhängig in der Parietalregion, was sich in Verbindung mit den umfangreichsten Artefaktvolumina in dieser Region in einer deutlichen Korrelation zeigte. Dabei fiel auf, dass PVE-korrigierte Werte in erheblicherem Maße von

fehlerhaften Segmentierungen betroffen waren. Zudem stellten wir ab einem Artefaktvolumen von >5,06 cm^3 (≙1500 Voxel) beträchtlich höhere Störungen fest. Eine Root-Mean-Square-Error (RMSE)-Berechnung der Untergruppen </≥1500 Artefaktvoxel zeigte für die 57 Segmentierungen <1500 Artefaktvoxel, unabhängig von Tracer und der Verwendung der PVEC, RMSE-[%] Werte von unter 1,5%. Im Gegensatz dazu standen die neun Segmentierungen ≥1500 Artefaktvoxel mit RMSE-[%] Werten bis zu 20%. In den longitudinalen Untersuchungen ergaben sich für die Bilddaten mit <1500 Artefaktvoxeln in der am stärksten beeinträchtigten Parietalregion äußerst geringe durchschnittliche SUV-Verfälschungen von weniger als 0,2% und dies unabhängig von Tracer und PVEC. Im Kontrast dazu stand die kleine Anzahl an Segmentierungen mit ≥1500 Artefaktvoxeln, welche den höchsten Einfluss auf die PET-Quantifizierung hatte, mit SUV-Abweichungen von bis zu 36% in den PVE-korrigierten [^{18}F]-AV45 PET-Daten der Parietalregion. Im zeitlichen Verlauf waren erneut PVE-korrigierte Werte stärker von Artefakten beeinträchtigt als ihre unkorrigierten Pendants. Auf Grund der immensen Verfälschung durch große Artefaktvolumina haben wir die Möglichkeit eines vereinfachten Maskierungsprozesses eruiert. Hierfür wurde nach Entfernung der Artefakte aus den VOIs aller 238 Bilddaten ein Atlas generiert (Watanabe, Andersen et al. 2001). Daraus erstellten wir eine 100%- und eine 50%-Voxelmaske. Erstgenannte enthielt sämtliche Voxel jedes VOI-Sets. Wohingegen die 50%-Voxelmaske die Voxel von mindestens der Hälfte der VOI-Sets enthielt. Das Template wurde bei den neun Bilddaten mit mehr als 1500 Artefaktvoxeln angewendet. Bei der 100%-Voxelmaske kam es nur zur marginalen Reduzierung der Artefaktvolumina. Wohingegen bei der 50%-Voxelmaske die meist betroffene Parietalregion

deutlich an Artefakten reduziert wurde, jedoch im Gegenzug korrekt segmentierte GM in der Frontalregion ebenfalls durch die Maskierung entfernt wurde. Zudem reduzierte sich in der Parietalregion der RMSE-[%] lediglich um etwa 50% unter Verwendung der 50%-Voxelmaske, unabhängig von Tracer und PVEC.

3. Zusammenfassung

Ziel der ersten Arbeit war die systematische Analyse des Einflusses unterschiedlicher Referenzregionen und einer Zielvolumen-basierten PVEC auf Amyloid PET-Quantifizierungen bei Alzheimer-Erkrankung und deren Auswirkung auf die Trennschärfe zwischen den Diagnosegruppen AD, MCI und HC (Brendel, Hogenauer et al. 2015). Insbesondere die Differenzierung letztgenannter Gruppen wird für zukünftige Therapieerprobungen entscheidend sein. Aus diesem Grund galt es auch, die Effekte unterschiedlicher Referenzregionen und der PVEC auf die longitudinale Amyloidlast zu untersuchen.

Dafür wurden Bilddaten eines ADNI-Kollektivs, bestehend aus [^{18}F]-AV45 PET-Daten (Amyloid PET) und T1-gewichteten MRT-Bildern, von 962 Probanden zur Baseline und von 258 Probanden im zweijährigen longitudinalen Verlauf untersucht. Aus dem Frontal-, Parietal-, Temporalcortex sowie precunealem/posteriorem cingulärem Cortex wurde eine COMP-VOI generiert. Für die Analysen wurden SUVs der vollständigen COMP-VOI (COMP-FULL-VOI) sowie der segmentierten GM ohne (COMP-GM-VOI) und mit PVEC (COMP-PVEC-GM-VOI) verwendet. Die Normalisierung der COMP-SUV-Werte erfolgte mit dem Cerebellum ($SUVR_{CBL}$), Hirnstamm ($SUVR_{BST}$) oder weißer Substanz ($SUVR_{WM}$). Die Mittelwerte von SUV, SUVR und der Einfluss einer PVEC wurden zwischen den Diagnosegruppen HC (N=316), MCI (N=483) und AD (N=163) in den Baselinedaten verglichen, die Trennschärfe zwischen den Diagnosegruppen mit Hilfe einer ROC-Analyse berechnet und als Area-Under-the-Curve (AUC) dargestellt. Die Konsequenz der Maßnahmen für die Bewertung, der sich

offensichtlich über die Zeit verändernden Amyloidlast, konnte schließlich mit Hilfe der longitudinalen Aß-Daten ermittelt werden.

In dieser Veröffentlichung erwies sich die WM zwischen den Diagnosegruppen als die beständigste Referenz. Die Anwendung der PVEC reduzierte die SUVs der COMP-GM-VOI-Konfiguration um -18% (HC) und -10% (MCI), während es bei den AD zu einem Anstieg von 7% kam. Der Einsatz der PVEC führte somit zu einer verbesserten Trennschärfe zwischen den Diagnosegruppen und ergab die höchsten AUC-Werte in Verbindung mit WM für HC/AD (0,907) und BST bei HC/MCI (0,658). Unabhängig von der Diagnose zeichnete sich unter Anwendung der PVEC der longitudinale Amyloidanstieg, mit Werten von bis zu 4,6%, am deutlichsten ab. Im zeitlichen Verlauf zeigte die WM als Referenz zudem die geringste Variabilität zwischen den Probanden. Im Rahmen dieser Studie mussten einige Datensätze wegen korrupter Segmentierungen der MRT-Daten von den weiteren Analysen ausgeschlossen werden (Baseline: N=56 (5,5%)/Follow-up: N=20 (7,2%)).

Daher galt es in der zweiten Arbeit, die Auswirkung von Segmentierungsartefakten auf die Quantifizierung von [^{18}F]-AV45 und [^{18}F]-FDG PET-Daten mit und ohne PVEC zu untersuchen (Hogenauer, Brendel et al. 2016). Dies erfolgte an 119 Probanden der ADNI-Kohorte, die T1-gewichtete MRT-Datensätze und [^{18}F]-AV45/[^{18}F]-FDG PET-Daten zur Baseline und im zweijährigen Verlauf vorweisen konnten. MRT-Daten mit Segmentierungsartefakten wurden mit Hilfe der zugehörigen [^{18}F]-FDG PET-Binärbilder maskiert. Um den Einfluss der Artefakte auf PET-Analysen zu quantifizieren, wurden sechs verschiedene VOIs definiert. Für jede dieser VOIs wurde für beide Tracer der prozentuale Fehler und der RMSE sowohl

für unkorrigierte als auch PVE-korrigierte SUVs berechnet und die Artefaktvolumina ermittelt. Die Effekte auf longitudinale PET-Daten wurden ebenfalls erhoben.

Im Zuge der Artefaktuntersuchungen stellten sich für beide PET-Tracer die prozentualen Fehler in Abhängigkeit der Größe des Artefaktvolumens dar und es konnten Korrelationskoeffizienten von bis zu 0,99 ermittelt werden. PVE-korrigierte SUVs waren deutlich stärker betroffen als es ihre unkorrigierten Pendants waren. Die parietale Region erwies sich zudem am häufigsten von Artefakten betroffen (30,4%). Eine kleine Untergruppe mit äußerst großen Artefakten (≥1500 Voxel, ≙5,06 cm^3) führte zu den größten Verfälschungen der PET-Quantifizierungen, was sich ebenfalls in den longitudinalen Auswertungen mit SUV-Abweichungen von bis zu 36% fortsetzte.

Insgesamt konnte mit dieser Dissertation gezeigt werden, dass die Zielvolumen-basierte PVEC und die Verwendung von WM oder BST als Referenz die Trennschärfe von [^{18}F]-AV45 PET-Daten im größeren Maße verbesserten als eine Kombination mit dem CBL. Die Nachweisempfindlichkeit von zunehmendem Aß im zeitlichen Verlauf wurde durch PVEC und WM als Referenzregion optimiert. Um die sichtlichen Vorteile dieser Maßnahmen nutzen zu können, müssen jedoch ausgedehnte Segmentierungsartefakte erkannt und beseitigt werden, um Verfälschungen in PET-Analysen zu vermeiden.

4. Summary

The aim of the first publication was a systematical analysis of the impact of different reference regions and VOI-based PVEC on the quantitation of amyloid PET results in Alzheimer's disease and their effect on the discriminatory power between diagnostic groups of AD, MCI and HC (Brendel, Hogenauer et al. 2015). In particular, the differentiation of the latter groups is essential for monitoring therapeutic trials in future. For this reason, it was also necessary to investigate the effects of different reference regions and PVEC on longitudinal amyloid load.

For this purpose, image data of an ADNI-cohort consisting of [^{18}F]-AV45 PET data (amyloid PET) and T1-weighted MRIs of 962 subjects to the baseline and 258 subjects in the two-year longitudinal course were examined. A COMP-VOI was generated from frontal, parietal and temporal cortex as well as precuneal/posterior cingulate cortex. For the analyses, SUVs of entire COMP-VOI (COMP-FULL-VOI) as well as segmented GM, uncorrected (COMP-GM-VOI) and PVE-corrected (COMP-PVEC-GM-VOI), were appropriated. The COMP-SUV values were normalized with the cerebellum ($SUVR_{CBL}$), brainstem ($SUVR_{BST}$) or white matter ($SUVR_{WM}$). The average values of SUV, SUVR and the influence of PVEC were compared at baseline between diagnostic groups of the HC (N=316), MCI (N=483) and AD (N=163) and discriminatory power between diagnostic groups were calculated by ROC-analyses and illustrated as AUC. Implications of quantitation procedures on obvious changes in amyloid load over time were finally determined by longitudinal Aß-data.

In this publication, WM proved to be the most consistent reference region between diagnosis groups. After PVEC application SUVs of COMP-GM-VOI

configuration decreased by -18% (HC) and -10% (MCI), while the AD increased by 7%. Consequently, the use of PVEC resulted in improved discriminatory power between diagnosis groups and resulted in the highest AUC values in conjunction with WM for HC/AD (0.907) and BST for HC/MCI (0.658). Regardless of the diagnosis, the longitudinal increase in amyloid grew the most with application of PVEC, with values up to 4.6%. Furthermore, WM revealed the lowest variability between the subjects as a reference in chronological sequence. In the context of this study some data sets had to be excluded from further analyses due to corrupt MR segmentations (Baseline: N=56 (5.5%)/Follow-up: N=20 (7.2%)).

Therefore, the second publication examined the impact of segmentation artifacts on the quantitation of uncorrected and PVE-corrected [^{18}F]-AV45 and [^{18}F]-FDG PET data (Hogenauer, Brendel et al. 2016). This was carried out at 119 subjects from the ADNI cohort, who could demonstrate T1-weighted MRI data sets and [^{18}F]-AV45/[^{18}F]-FDG PET data at baseline and at two-year follow-up. MRI data with segmentation artifacts were masked with the corresponding [^{18}F]-FDG PET binary images. Six different VOIs were defined to quantify the impact of artifacts on PET analyses. For each of these VOIs, the percentage error and RMSE for both tracers were calculated for uncorrected as well as PVE-corrected SUVs and the artifact volumes were determined. The effects on longitudinal PET data were also recorded.

In the course of the artifact assessment, the percentage errors were found for both PET tracers depending on the size of the artifact volume and correlation coefficients of up to 0.99 could be determined. PVE-corrected SUVs were significantly more affected than their uncorrected counterparts.

The parietal cortex region was also most frequently affected by artifacts (30.4%). A small subgroup with extremely large artifacts (≥1500 voxel, ≙5.06 cm^3) led to the greatest distortions in PET quantitation, which also continued in the longitudinal evaluations with SUV deviations up to 36%.

Overall, this doctoral thesis showed that the VOI-based PVEC and the use of WM or BST as reference improved the discriminatory power of [^{18}F]-AV45 PET data to a greater extent than a combination with the CBL. The detection sensitivity of increasing Aß over time was optimized by PVEC and WM as a reference region. However, to make use of the obvious advantages of this measure, extensive segmentation artifacts need to be identified and eliminated in order to avoid distortions in PET analyses.

5. Abkürzungsverzeichnis

Aß	=	ß-Amyloid
AD	=	Alzheimer-Patienten
ADNI	=	Alzheimer`s Disease Neuroimaging Initiative
APP	=	Amyloid-Vorläufer-Protein
AUC	=	Area-Under-the-Curve
BST	=	Hirnstamm
CBL	=	Cerebellum
COMP VOI	=	kombinierte Volume-Of-Interest
CSF	=	Liquor
FWHM	=	Full-Width-at-Half-Maximum
GM	=	graue Hirnsubstanz
GTM	=	Geometric-Transfer-Matrix
HC	=	gesunde Kontrollpersonen
MCI	=	leicht kognitiv Beeinträchtigte
MNI	=	Montreal Neurological Institute
MRT	=	Magnet-Resonanz-Tomographie
PET	=	Positronen-Emissions-Tomographie
PSF	=	Point-Spread Function
PVE	=	Partialvolumeneffekt
PVEC	=	Partialvolumeneffektkorrektur
RMSE	=	Root-Mean-Square-Error
ROC	=	Receiver-Operating-Characteristic
SUV	=	Standard-Uptake-Value
SUVR	=	Standard-Uptake-Value-Ratio
VOI	=	Volume-Of-Interest
WM	=	weiße Hirnsubstanz
[^{11}C]-PiB	=	[^{11}C]-Pittsburgh compound B
[^{18}F]-AV45	=	[^{18}F]-Florbetapir
[^{18}F]-FDG	=	[^{18}F]-Fluorodesoxyglucose

6. Literaturverzeichnis

Alzheimer's Disease International (2015). "World Alzheimer Report 2015, The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends." Online: Alzheimer's Disease International: https://www.alz.co.uk/research/world-report-2015 ; [PDF-Datei]: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [Stand: 28.07.2016].

Ashburner, J. and K. J. Friston (2005). "Unified segmentation." Neuroimage **26**(3): 839-851.

Bickel, H. (2016). "Infoblatt1: Die Häufigkeit von Demenzerkrankungen." Online: Deutsche Alzheimer Gesellschaft e. V.: https://www.deutsche-alzheimer.de/unser-service/informationsblaetter-downloads.html ; [PDF-Datei]: https://www.deutsche-alzheimer.de/fileadmin/alz/pdf/factsheets/infoblatt1_haeufigkeit_demenzerkrankungen_dalzg.pdf [Stand: 28.07.2016].

Bohnen, N. I., D. S. Djang, et al. (2012). "Effectiveness and safety of 18F-FDG PET in the evaluation of dementia: a review of the recent literature." J Nucl Med **53**(1): 59-71.

Borghammer, P., P. Cumming, et al. (2009). "Subcortical elevation of metabolism in Parkinson's disease--a critical reappraisal in the context of global mean normalization." Neuroimage **47**(4): 1514-1521.

Brendel, M., A. Delker, et al. (2014). "Impact of partial volume effect correction on cerebral beta-amyloid imaging in APP-Swe mice using [(18)F]-florbetaben PET." Neuroimage **84**: 843-853.

Brendel, M., M. Hogenauer, et al. (2015). "Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction." Neuroimage **108**: 450-459.

Brookmeyer, R., E. Johnson, et al. (2007). "Forecasting the global burden of Alzheimer's disease." Alzheimers Dement **3**(3): 186-191.

Clark, C. M., M. J. Pontecorvo, et al. (2012). "Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study." Lancet Neurol **11**(8): 669-678.

Corder, E. H., A. M. Saunders, et al. (1994). "Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease." Nat Genet **7**(2): 180-184.

de Leon, M. J., A. Convit, et al. (2001). "Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]fluoro-2-deoxy-D-glucose/poitron-emission tomography (FDG/PET)." Proc Natl Acad Sci U S A **98**(19): 10966-10971.

Dukart, J., K. Mueller, et al. (2010). "Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies." Neuroimage **49**(2): 1490-1495.

Dukart, J., R. Perneczky, et al. (2013). "Reference cluster normalization improves detection of frontotemporal lobar degeneration by means of FDG-PET." PLoS One **8**(2): e55415.

Duyckaerts, C., B. Delatour, et al. (2009). "Classification and basic pathology of Alzheimer disease." Acta Neuropathol **118**(1): 5-36.

Engler, H., A. Forsberg, et al. (2006). "Two-year follow-up of amyloid deposition in patients with Alzheimer's disease." Brain **129**(Pt 11): 2856-2866.

Erlandsson, K., I. Buvat, et al. (2012). "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology." Phys Med Biol **57**(21): R119-159.

Gauthier, S., M. Albert, et al. (2016). "Why has therapy development for dementia failed in the last two decades?" Alzheimers Dement **12**(1): 60-64.

Goedert, M. and R. Jakes (2005). "Mutations causing neurodegenerative tauopathies." Biochim Biophys Acta **1739**(2-3): 240-250.

Gotz, J., F. Chen, et al. (2001). "Formation of neurofibrillary tangles in P301l tau transgenic mice induced by Abeta 42 fibrils." Science **293**(5534): 1491-1495.

Hammers, A., R. Allom, et al. (2003). "Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe." Hum Brain Mapp **19**(4): 224-247.

Hampel, H., D. Prvulovic, et al. (2011). "The future of Alzheimer's disease: the next 10 years." Prog Neurobiol **95**(4): 718-728.

Hogenauer, M., M. Brendel, et al. (2016). "Impact of MRI-based Segmentation Artifacts on Amyloid- and FDG-PET Quantitation." Curr Alzheimer Res **13**(5): 597-607.

Hugo, J. and M. Ganguli (2014). "Dementia and cognitive impairment: epidemiology, diagnosis, and treatment." Clin Geriatr Med **30**(3): 421-442.

Hutton, B. F. and A. Osiecki (1998). "Correction of partial volume effects in myocardial SPECT." J Nucl Cardiol **5**(4): 402-413.

Jack, C. R., Jr., D. S. Knopman, et al. (2010). "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade." Lancet Neurol **9**(1): 119-128.

Jack, C. R., Jr., T. M. Therneau, et al. (2016). "Transition rates between amyloid and neurodegeneration biomarker states and to dementia: a population-based, longitudinal cohort study." Lancet Neurol **15**(1): 56-64.

Jagust, W., A. Gitcho, et al. (2006). "Brain imaging evidence of preclinical Alzheimer's disease in normal aging." Ann Neurol **59**(4): 673-681.

Johnson, K. A., N. C. Fox, et al. (2012). "Brain imaging in Alzheimer disease." Cold Spring Harb Perspect Med **2**(4): a006213.

Johnson, K. A., R. A. Sperling, et al. (2013). "Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging." Alzheimers Dement **9**(5 Suppl): S72-83.

Joshi, A. D., M. J. Pontecorvo, et al. (2012). "Performance characteristics of amyloid PET with florbetapir F 18 in patients with alzheimer's disease and cognitively normal subjects." J Nucl Med **53**(3): 378-384.

Leung, K. K., J. Barnes, et al. (2011). "Brain MAPS: an automated, accurate and robust brain extraction technique using a template library." Neuroimage **55**(3): 1091-1108.

Mikheev, A., G. Nevsky, et al. (2008). "Fully automatic segmentation of the brain from T1-weighted MRI using Bridge Burner algorithm." J Magn Reson Imaging **27**(6): 1235-1241.

Minoshima, S., K. A. Frey, et al. (1995). "Preserved pontine glucose metabolism in Alzheimer disease: a reference region for functional brain image (PET) analysis." J Comput Assist Tomogr **19**(4): 541-547.

Mueller, S. G., M. W. Weiner, et al. (2005). "The Alzheimer's disease neuroimaging initiative." Neuroimaging Clin N Am **15**(4): 869-877, xi-xii.

Mueller, S. G., M. W. Weiner, et al. (2005). "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)." Alzheimers Dement **1**(1): 55-66.

Muller-Gartner, H. W., J. M. Links, et al. (1992). "Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects." J Cereb Blood Flow Metab **12**(4): 571-583.

Nasrallah, I. M. and D. A. Wolk (2014). "Multimodality imaging of Alzheimer disease and other neurodegenerative dementias." J Nucl Med **55**(12): 2003-2011.

Nordberg, A. (2011). "Molecular imaging in Alzheimer's disease: new perspectives on biomarkers for early diagnosis and drug development." Alzheimers Res Ther **3**(6): 34.

Poirier, J., J. Davignon, et al. (1993). "Apolipoprotein E polymorphism and Alzheimer's disease." Lancet **342**(8873): 697-699.

Querfurth, H. W. and F. M. LaFerla (2010). "Alzheimer's disease." N Engl J Med **362**(4): 329-344.

Rousset, O. G., D. L. Collins, et al. (2008). "Design and implementation of an automated partial volume correction in PET: application to dopamine receptor quantification in the normal human striatum." J Nucl Med **49**(7): 1097-1106.

Rousset, O. G., P. Deep, et al. (2000). "Effect of partial volume correction on estimates of the influx and cerebral metabolism of 6-[(18)F]fluoro-L-dopa studied with PET in normal control and Parkinson's disease subjects." Synapse **37**(2): 81-89.

Rousset, O. G., Y. Ma, et al. (1998). "Correction for partial volume effects in PET: principle and validation." J Nucl Med **39**(5): 904-911.

Segobin, S., R. La Joie, et al. (2015). "FDG-PET Contributions to the Pathophysiology of Memory Impairment." Neuropsychol Rev **25**(3): 326-355.

Segonne, F., A. M. Dale, et al. (2004). "A hybrid approach to the skull stripping problem in MRI." Neuroimage **22**(3): 1060-1075.

Smith, S. M. (2002). "Fast robust automated brain extraction." Hum Brain Mapp **17**(3): 143-155.

Thomas, B. A., K. Erlandsson, et al. (2011). "The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease." Eur J Nucl Med Mol Imaging **38**(6): 1104-1119.

Thompson, P. M., K. M. Hayashi, et al. (2003). "Dynamics of gray matter loss in Alzheimer's disease." J Neurosci **23**(3): 994-1005.

van Berckel, B. N., R. Ossenkoppele, et al. (2013). "Longitudinal amyloid imaging using 11C-PiB: methodologic considerations." J Nucl Med **54**(9): 1570-1576.

Villemagne, V. L., S. Burnham, et al. (2013). "Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study." Lancet Neurol **12**(4): 357-367.

Watanabe, H., F. Andersen, et al. (2001). "MR-based statistical atlas of the Gottingen minipig brain." Neuroimage **14**(5): 1089-1096.

Weiner, M. W., P. S. Aisen, et al. (2010). "The Alzheimer's disease neuroimaging initiative: progress report and future plans." Alzheimers Dement **6**(3): 202-211 e207.

Weiner, M. W., D. P. Veitch, et al. (2013). "The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception." Alzheimers Dement **9**(5): e111-194.

Winblad, B., P. Amouyel, et al. (2016). "Defeating Alzheimer's disease and other dementias: a priority for European science and society." Lancet Neurol **15**(5): 455-532.

Wolk, D. A., Z. Zhang, et al. (2012). "Amyloid imaging in Alzheimer's disease: comparison of florbetapir and Pittsburgh compound-B positron emission tomography." J Neurol Neurosurg Psychiatry **83**(9): 923-926.

Yakushev, I., C. Landvogt, et al. (2008). "Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18." Psychiatry Res **164**(2): 143-153.

7. Danksagung

An erster Stelle gilt mein Dank meinem Doktorvater Herrn Prof. Dr. med. Axel Rominger für das Anvertrauen dieser hochinteressanten Thematik. Für die exzellente Betreuung, das stets offene Ohr und die konstruktiven Impulse sowie die außerordentlich angenehme Atmosphäre in jeder Bearbeitungsphase, von der Einführung über die Erarbeitung und Auswertung der Ergebnisse bis hin zur Fertigstellung dieser Arbeit, spreche ich meinen besonderen Dank aus.

Mein weiterer Dank gilt Herrn Dr. med. Matthias Brendel für die erstklassige Einführung in die Thematik und in das Programm PMOD. Die zahlreichen anregenden Diskussionen und Ratschläge, seine stetige Erreichbarkeit und nicht zuletzt die ausgesprochen angenehme Zusammenarbeit während der gesamten Datenaufbereitung und -auswertung haben maßgeblich zum Gelingen dieser Dissertation beigetragen.

Weiterhin möchte ich mich bei Herrn Prof. Dr. med. Peter Bartenstein, Direktor der Klinik und Poliklinik für Nuklearmedizin der Universität München, bedanken, der mir ermöglichte, in seiner Abteilung diese Thematik zu bearbeiten.

Ein großer Dank gebührt ebenso der gesamten Arbeitsgruppe, insbesondere den Physikern, Guido Böning und Andreas Delker, für die äußerst kollegiale Arbeitsatmosphäre und die stetige Hilfsbereitschaft sowie meinen Mitdoktoranden, für die vielen kurzweiligen, interessanten und immer in Erinnerung bleibenden Stunden im Doktorandenraum.

An dieser Stelle ein herzlicher Dank meinen Eltern, für die Ermöglichung des Studiums, und meinem Bruder, die mir alle viel Geduld und stets ein offenes Ohr für meine Gedanken während dieser Dissertation entgegenbrachten.

NeuroImage 108 (2015) 450–459

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improved longitudinal [^{18}F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction

Matthias Brendel [a], Marcus Högenauer [a], Andreas Delker [a], Julia Sauerbeck [a], Peter Bartenstein [a], John Seibyl [b], Axel Rominger [a,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Dept. of Nuclear Medicine, University of Munich, Germany
[b] MNI, New Haven, USA

ARTICLE INFO

Article history:
Accepted 26 November 2014
Available online 4 December 2014

Keywords:
^{18}F-AV45-PET
Alzheimer's disease
Reference region
Partial volume effect correction

ABSTRACT

Amyloid positron-emission-tomography (PET) offers an important research and diagnostic tool for investigating Alzheimer's disease (AD). The majority of amyloid PET studies have used the cerebellum as a reference region, and clinical studies have not accounted for atrophy-based partial volume effects (PVE). Longitudinal studies using cerebellum as reference tissue have revealed only small mean increases and high inter-subject variability in amyloid binding. We aimed to test the effects of different reference regions and PVE-correction (PVEC) on the discriminatory power and longitudinal performance of amyloid PET.

We analyzed [^{18}F]-AV45 PET and T1-weighted MRI data of 962 subjects at baseline and two-year follow-up data of 258 subjects. Cortical composite volume-of-interest (VOI) values (COMP) for tracer uptake were generated using either full brain atlas VOIs, gray matter segmented VOIs or gray matter segmented VOIs after VOI-based PVEC. Standard-uptake-value ratios (SUVR) were calculated by scaling the COMP values to uptake in cerebellum ($SUVR_{CBL}$), brainstem ($SUVR_{BST}$) or white matter ($SUVR_{WM}$). Mean SUV, SUVR, and changes after PVEC were compared at baseline between diagnostic groups of healthy controls (HC; N = 316), mild cognitive impairment (MCI; N = 483) and AD (N = 163). Receiver operating characteristics (ROC) were calculated for the discriminations between HC, MCI and AD, and expressed as area under the curve (AUC). Finally, the longitudinal [^{18}F]-AV45-PET data were used to analyze the impact of quantitation procedures on apparent changes in amyloid load over time.

Reference region SUV was most constant between diagnosis groups for the white matter. PVEC led to decreases of COMP-SUV in HC (−18%) and MCI (−10%), but increases in AD (+7%). Highest AUCs were found when using PVEC with white matter scaling for the contrast between HC/AD (0.907) or with brainstem scaling for the contrast between HC/MCI (0.658). Longitudinal increases were greatest in all diagnosis groups with application of PVEC, and inter-subject variability was lowest for the white matter reference.

Thus, discriminatory power of [^{18}F]-AV45-PET was improved by use of a VOI-based PVEC and white matter or brainstem rather than cerebellum reference region. Detection of longitudinal amyloid increases was optimized with PVEC and white matter reference tissue.

Abbreviations: Aß, ß-amyloid; AD, Alzheimer's disease; ADAS-cog, Alzheimer's disease assessment scale – cognitive subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; AUC, area under the curve; BST, brainstem; CBL, cerebellum; COMP, combined region of frontal, parietal, temporal and posterior cingulate cortices; CSF, cerebrospinal fluid; FAD, familiar Alzheimer's disease; FDA, Food and Drug Administration; FDG, fluorodeoxyglucose; FULL, full atlas VOIs; GM, gray matter; HC, cognitively healthy; MANOVA, multivariate analysis of covariance; MCI, mild cognitive impairment; MMSE, mini mental state examination; MNI, Montreal Neurological Institute; MRI, magnetic resonance imaging; NIA, National Institute on Aging; NIBIB, National Institute of Biomedical Imaging and Bioengineering; PET, positron emission tomography; PVE, partial volume effect; PVEC, partial volume effect correction; R, Pearson's coefficient of correlation; REF, reference configuration from tracer validation; ROC, receiver operating characteristics; SAD, sporadic Alzheimer's disease; SUV, standard uptake value; SUVR, standard uptake value ratio; T1w, T1 weighted; VOI, volume of interest; WM, white matter; FWHM, Full-width-at-half-maximum.

* Corresponding author at: Department of Nuclear Medicine, University of Munich, Germany. Fax: +49 89 4400 77646.
E-mail address: axel.rominger@med.uni-muenchen.de (A. Rominger).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

http://dx.doi.org/10.1016/j.neuroimage.2014.11.055
1053-8119/

Introduction

Alzheimer's disease (AD) is the most common form of dementia; its incidence increases exponentially as a function of age, which is imposing an onerous burden on health care systems in societies with aging populations (Ziegler-Graham et al., 2008). Neurofibrillary tangles and amyloid plaques together comprise the hallmark neuropathology of AD (Braak and Braak, 1991). Elevated brain amyloid burden is now clearly associated with cognitive decline in the healthy elderly (HC) (Lim et al., 2012) and in cases of mild cognitive impairment (MCI) (Lim et al., 2014). Amyloid PET offers a feasible tool for the early detection of brain amyloidosis, and the recent development of fluorine-18 labeled amyloid radioligands such as [^{18}F]-AV45 has made this technique available to PET centers lacking an on-site cyclotron/radiochemistry facility.

In clinical PET practice, Aβ-positivity and -negativity are visually assessed with good inter- and intra-reader agreement (Clark et al., 2012). However, a semiquantitative approach is better suited especially to the requirements of longitudinal clinical trials of amyloidosis progression and treatment. The issue of defining an optimal reference region has been extensively discussed for normalization of [^{18}F]-fluorodeoxyglucose-(FDG) PET relative to cerebellum, pons/brainstem, global mean, or a reference cluster (Bohnen et al., 2012; Borghammer et al., 2009; Dukart et al., 2013; Yakushev et al., 2008).

In PET imaging with [^{18}F]-AV45- and [^{11}C]-PiB, the entire cerebellum and the cerebellar gray matter (GM) have emerged as the most widely used reference regions for quantitation of amyloid burden (Weiner et al., 2013). However, a recent longitudinal [^{11}C]-PiB PET study of mild cognitive impairment (MCI) and AD showed high inter-subject variability based on a cerebellar GM reference (van Berckel et al., 2013). Furthermore, amyloid PET results are potentially biased by partial volume effects (PVE), which have a considerable impact in patients with pronounced atrophy (Thomas et al., 2011), which is particularly problematic in longitudinal studies.

Given these considerations, we aimed to compare systematically the quantitation of [^{18}F]-AV45-PET results for different reference regions, using as our material the Alzheimer's Disease Neuroimaging Initiative (ADNI)-dataset, which includes more than 1000 amyloid PET cases. Furthermore, we set about to investigate the impact of a volume-of-interest (VOI)-based partial volume effect correction (PVEC) on the semiquantitative analyses. Receiver operating characteristics (ROC) were obtained for the baseline discrimination of HC from MCI and AD cases in order to identify the most sensitive amyloid-PET analysis. Finally, two-year longitudinal [^{18}F]-AV45-PET data from 258 patients used to test the impact of the above factors on apparent changes in amyloid load with time.

Materials and methods

Alzheimer's disease neuroimaging initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California — San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data from ADNI-GO/-2 were included in this work. Pre-processed brain [^{18}F]-AV45-PET images and temporally corresponding T1-weighted MPRAGE images were downloaded from the ADNI database as available on Jan 16th, 2014.

Patient selection and study design

1018 subjects from ADNI-GO and ADNI-2 who had undergone brain [^{18}F]-AV45-PET and T1 weighted MPRAGE (T1w) MRI at their study baseline were included in this investigation. Clinical diagnoses as provided by the ADNI database at the time of PET-imaging were: 336 HC, 508 MCI and 174 AD. A subgroup of 278 subjects (94 HC, 158 MCI and 26 AD) who underwent a second [^{18}F]-AV45-PET and T1w MRI at two-year follow-up was used for longitudinal analyses.

Image data

ADNI [^{18}F]-AV45-PET acquisition and pre-processing

The [^{18}F]-AV45-PET images had been acquired using Siemens, GE and Philips PET scanners according to a standard dynamic 50–70 min protocol following the intravenous injection of 370 ± 37 MBq of [^{18}F]-AV45. Data were corrected for both scatter and measured attenuation, which was determined using the CT scan for PET/CT scanners, or a transmission scan with [^{68}Ge] or [^{137}Cs] rotating rod sources for PET-only scanners. Images were reconstructed using scanner-specific algorithms, and sent to the University of Michigan, where they were reviewed for artifacts and transmitted to the Laboratory of NeuroImaging (LONI) for storage.

Downloaded [^{18}F]-AV45-PET images in DICOM format had been pre-processed in four steps: 1) motion correction by co-registration of single five minute frames; 2) time frame averaging (50–70 min p.i.); 3) co-registration of longitudinal data to the baseline scan and reorientation in a standardized 160 × 160 × 96 matrix with 1.5 mm cubic voxels; 4) smoothing with a scanner-specific filter function to an isotropic resolution of 8 mm. Further details are provided in Supplement 1.

ADNI MRI acquisition and pre-processing

T1-weighted MRI scans had been acquired using Siemens, GE or Philips MRI scanners according to a standard protocol (Jack et al., 2008) involving acquisitions of two 3-D MPRAGE imaging sequences per subject. Of the two images acquired per subject and time-point, the ADNI quality assurance team selected the better image for pre-processing, based on the presence and severity of commonly occurring image artifacts.

MRI preprocessing involved: 1) application of a scanner-specific correction for gradient nonlinearity distortion (Gradwarp) (Jovicich et al., 2006); 2) correction for image intensity non-uniformity (B1) (Jack et al., 2008); 3) histogram peak sharpening algorithm for bias field correction (N3) (Sled et al., 1998.); 4) application of spatial scaling factors obtained by phantom measurements. For images acquired on Philips scanners, B1 correction was already implemented, and the gradient systems with this instrument tended to be linear (Jack et al., 2008).

PET analyses

Co-registration and quality control

All procedures were performed according to an automatic protocol using the PMOD PNEURO tool (V. 3.5 PMOD Technologies, Zürich): PET images were rigidly co-registered to the corresponding MRI to calculate a linear transformation (PET-2-MRI). Individual MRI images were nonlinearly co-registered to the standard Montreal Neurological Institute (MNI) space MRI template (MRI-2-MNI), while PET-2-MRI and MRI-2-MNI transformations were used to transform PET images into the MNI space.

T1-weighted MR images were segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) (Ashburner and Friston, 2005) to generate a total of 83 individual brain VOIs for each subject in the MNI space, according to the atlas of Hammers (Hammers et al., 2003). Hammers Atlas VOIs were also defined and assessed without segmentation (FULL). In addition to the Hammers VOIs, three white matter VOIs were defined and tested: 1) WM, for which the area of individual cortical GM VOIs was subtracted from entire brain volume, thus segmenting all white matter voxels included in the whole atlas VOI (this, separately for cortex and cerebellum). 2) WM-seg, for which the segmentation process was used to generate WM VOIs containing all voxels declared as white matter in the segmentation (separately for cortex and cerebellum). 3) CSO, a manually drawn bilateral VOI for the centrum semiovale, of volume 2.4 cm^3. More detailed illustrations of the different white matter volumes are given in Supplement 2 (Fig. S1). Results for WM are reported in the manuscript, while results for WM-seg and CSO are provided in Supplement 2 (Tables S1–S4).

Inverse PET-2-MNI transformations were used to resample VOIs from the Hammers atlas to the native PET space, in which VOI calculations were subsequently performed. GM volumes were assessed in MNI space to account for different brain sizes. Further details of the analysis procedure are provided in Supplement 3 (Fig. S2).

All images were visually checked for correct co-registration and appropriate segmentation. From a total of 1018 (baseline) PET images, 36 (3.5%) were identified as misregistered, and were consequently processed semi-automatically by inclusion of hot spot alignment or masking of extracerebral structures prior to final co-registration. A total of 56 (5.5%) MRI segmentations were found to be inadequate (due failed GM/WM/CSF identification or visible artifacts) when calculated using the standard settings. These 56 subjects were necessarily excluded from further analyses, as changing standard parameters could have biased the resulting VOI configuration. For the 278 follow-up PET images, eight (3%) had to be processed semi-automatically and 20 (7.2%) were excluded due to failed GM/WM/CSF identification or visible artifacts (Fig. 1). Details of the resulting study groups including demographics are provided in Table 1.

VOI-based PVEC

A VOI-based PVEC (Rousset et al., 1998, 2008) was performed in PET space for individual GM VOIs from the Hammers Atlas (again obtained through multiplication by the inverse PET-2-MNI transformation). Background and white matter regions were included in the correction algorithm to ensure that all voxels included in the atlas were assigned to a VOI. These VOIs were subsequently used for calculation of PVE-corrected values by the geometric transfer matrix method,

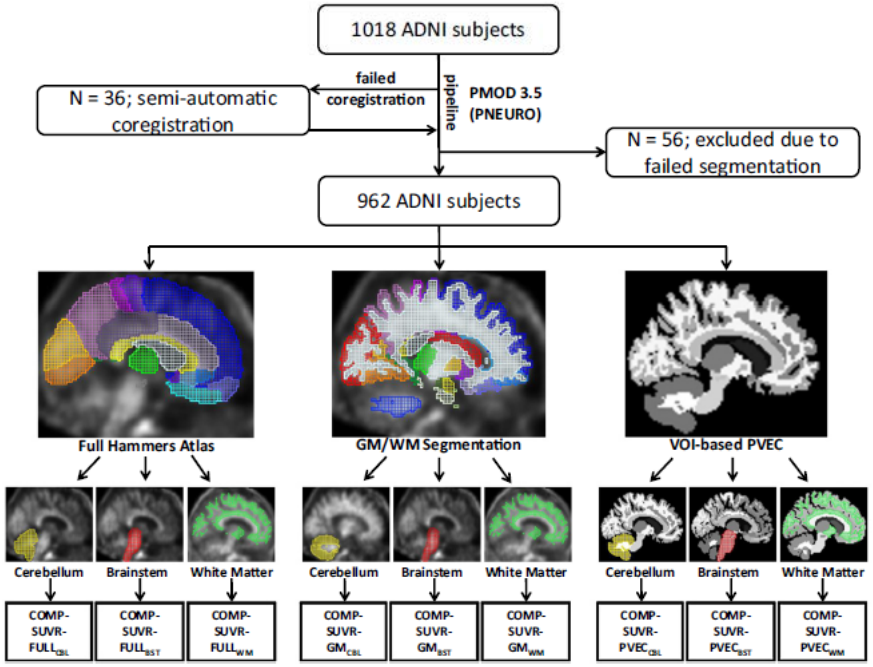


Fig. 1. Workflow from ADNI database to VOI-based analyses. All subjects receiving an [^{18}F]-AV45 PET scan and a T1 weighted MRI at baseline were processed using the full-automatic PNEURO (PMOD 3.5) pipeline. Subsequent visual inspection was followed either by manual coregistration when mis-registrations occurred, or by exclusion when segmentation errors or extensive artifacts were detected. Remaining subjects were analyzed using full atlas VOIs (Hammers et al., 2003), segmented GM VOIs and segmented GM VOIs after PVEC. Resulting VOI values were consecutively scaled by either the cerebellum (yellow), the brainstem (red) or the white matter (green). A tenth methodological configuration consisted of full atlas VOIs and scaling by the whole cerebellum and served as a reference (Johnson et al., 2013; Joshi et al., 2012).

Table 1
Demographics of healthy controls (HC), mild cognitive impaired patients (MCI) and patients suffering from Alzheimer's disease (AD) included in this study. Column 9 indicates subjects with available 2-year follow-up [^{18}F]-AV45-PET and MRI. * p< 0.05; ** p< 0.01; *** p< 0.001.

DX-group	Baseline N (Aß+,%)	Age (y ± SD)	Gender (f/m)	Education (y ± SD)	APOE ε4 (N of alleles ± SD)	ADAS-cog (0–70 ± SD)	MMSE (0–30 ± SD)	2-year follow up N (Aß+,%)
HC	316 (108, 32%)	74.6 ± 6.5	148 (47%) f/168 (53%) m	16.4 ± 2.6	0.31 ± 0.51	9.1 ± 4.6	29.0 ± 1.2	88 (26, 30%)
MCI	483 (267, 55%)	72.5 ± 7.9**	272 (57%) f/210 (43%) m	16.1 ± 2.7	0.56 ± 0.65*	14.7 ± 6.7***	28.0 ± 1.7**	148 (66, 45%)
AD	163 (143, 88%)	75.1 ± 7.7	99 (61%) f/64 (39%) m	16.0 ± 2.6	0.88 ± 0.71**	31.2 ± 9.5***	22.7 ± 3.2***	22 (16, 73%)

cited above. The gray matter threshold was set at 0.3 while a full-width-at-half-maximum (FWHM) of 8 mm was used, determined as the uniform spatial resolution after ADNI preprocessing. Both uncorrected and PVE-corrected results were generated (implemented in PNEURO) for individual GM VOIs. More details and visualization of the processing steps are provided in Supplement 3 (Fig. S2).

Assessment of VOI values

Following analyses of PET data were performed (Fig. 1):

a) Uncorrected and PVE-corrected standard uptake values (SUV) were calculated for a combined composite VOI (COMP) consisting of frontal, parietal, temporal cortex and precuneal/posterior cingulate gyrus. Cerebellum (CBL), brainstem (BST), and white matter (WM) were assessed as reference regions. COMP and CBL were assessed both as full atlas regions (FULL-VOI) and as segmented GM regions (GM-VOI). Additionally PVE-corrected values were obtained for the GM-VOI (PVEC-GM-VOI) of COMP and CBL.
b) Ten different composite VOI standard uptake value ratios (SUVR) were calculated: Composite FULL-VOI scaled either by (1) the cerebellar FULL-VOI (COMP-SUVR-FULL$_{CBL}$), (2) the brainstem VOI (COMP-SUVR-FULL$_{BST}$), or (3) the white matter VOI (COMP-SUVR-FULL$_{WM}$); composite GM-VOI scaled either by (4) the cerebellar GM-VOI (COMP-SUVR-GM$_{CBL}$), (5) the brainstem VOI (COMP-SUVR-GM$_{BST}$), or (6) the white matter VOI (COMP-SUVR-GM$_{WM}$); composite PVEC-GM-VOI scaled either by (7) the cerebellar PVEC-GM-VOI (COMP-SUVR-PVEC$_{CBL}$), (8) the brainstem VOI (COMP-SUVR-PVEC$_{BST}$), or (9) the white matter VOI (COMP-SUVR-PVEC$_{WM}$). The final configuration (10) consisted of the composite GM-VOI scaled by the cerebellar FULL-VOI (Johnson et al., 2013; Joshi et al., 2012), and served as validation of our analyses against this reference standard (COMP-SUVR-REF). Aβ-positivity/-negativity was obtained according to the COMP-SUVR-REF threshold of >/≤1.1 in ADNI database.
c) To ascertain the effect of VOI-based PVEC, changes in raw PVE-corrected VOI values (SUVs) were calculated relative to their uncorrected values (% change). GM volumes were extracted after segmentation to estimate the global GM atrophy.
d) Longitudinal changes after two years in [^{18}F]-AV45 uptake (Δ%-COMP-SUVR) were assessed for the ten composite SUVR according to Eq. (1):

$$\Delta\%-COMP-SUVR = \frac{[COMP-SUVR\,(2Y-Follow-Up)]-[COMP-SUVR\,(Baseline)]}{COMP-SUVR\,(Baseline)} \times 100\% \quad (1)$$

Statistics

a) Mean (± SD) SUV for the reference regions were calculated for HC, MCI and AD groups and compared using multivariate analysis of variance with subsequent Bonferroni correction (MANOVA).
b) Mean (± SD) COMP-SUVR were calculated for HC, MCI and AD groups according to all ten different evaluations listed above, and compared using MANOVA.
c) Mean (± SD) % changes after PVEC in GM-SUV of composite and cerebellum as well as GM volumes were compared between HC, MCI and AD groups using MANOVA.
d) ROCs were calculated for the discriminatory power between AD/HC, AD/MCI and MCI/HC for the ten composite SUVRs, using SPSS (version 22.0; SPSS, Chicago, IL). ROCs for the nine test configurations were contrasted against COMP-SUVR-REF using ROC_2_curves_tool version 6 (ACOMED statistic, Leipzig Germany) including the non-parametric approach of Delong (DeLong et al., 1988). [^{18}F]-AV45 cut-off values, sensitivities and specificities were calculated for revealing the best tradeoff between sensitivity and specificity for the discrimination between AD/HC, AD/MCI and MCI/HC.
e) Results from all 74 (33 HC, 39 MCI, 2 AD) non-amyloid accumulating subjects (those who were Aβ-negative in both [^{18}F]-AV45 PET assessments) were used to perform correlation analysis (Pearson's R) between baseline and follow-up COMP-SUVR for all ten different configurations with regard to longitudinal stability.
f) Mean (± SD) Δ%-COMP-SUVR were calculated by diagnosis groups for all ten COMP-SUVR evaluations. Paired t-tests were used for the comparison of different configurations against the COMP-SUVR-REF configuration. Mean (± SD) Δ%-volume of COMP and CBL were calculated by diagnosis groups and compared using MANOVA.

Different PET scanners (N = 17) of the multi-center study were included as a covariate for all group-wise calculations. For all statistical tests p-values < 0.05 were assigned to be significant.

Results

Demographics

See Table 1 for details of the study population. At baseline, 108/316 (32%) HC subjects, 267/483 (55%) MCI subjects and 143/163 (88%) AD subjects had a positive [^{18}F]-AV45 scan (SUVR > 1.10). From those subjects with an additional 2-year follow-up scan, 26/88 (30%) HC subjects, 66/148 (45%) MCI subjects and 16/22 (73%) AD subjects had a positive [^{18}F]-AV45 baseline scan. Nine converters to MCI and one converter to AD were observed in HC subjects. Five re-converters to HC and 12 converters to AD were found in the MCI group, while all AD subjects retained AD-positive status.

Comparison of reference region SUVs

All mean values (± SD) for the SUVs of reference regions by clinical group are provided in Table 2. Strong decreases in SUVs of AD subjects when compared to HC were observed for CBL-FULL-VOI (−4.3%;

Table 2
Mean (±SD) SUV for all reference regions (column 1) by diagnosis groups (columns 2–4). F-statistics and p-values for diagnosis related MANOVA are presented in columns 5–6. Significant differences against HC are indicated by: *p < 0.05; **p < 0.01; ***p < 0.001.

SUV region	HC	MCI	AD	F-statistic	p-value
CBL-FULL	1.09 ± 0.07	1.08 ± 0.07	1.04 ± 0.07***	23.8	<0.001
CBL-GM	1.03 ± 0.06	1.03 ± 0.05	1.02 ± 0.05*	3.5	<0.05
BST	1.63 ± 0.20	1.59 ± 0.19***	1.47 ± 0.18***	34.3	<0.001
WM	1.52 ± 0.23	1.54 ± 0.23	1.54 ± 0.25	1.1	n.s.

$p < 0.001$) and BST-VOI (−10.4%; $p < 0.001$), while a small but significant decrease was found for CBL-GM-VOI (−1.4%; $p < 0.05$) in this contrast. WM-VOI SUV did not show any significant differences between diagnosis groups.

Comparison of differently assessed composite SUVR

All mean values (± SD) for the ten differently assessed COMP-SUVRs by clinical group are provided in Table 3. Differences between MCI/HC and AD/HC were highly significant regardless of the assessment configuration used (all $p < 0.001$). The COMP-SUVRs were consistently lower when using the brainstem or the white matter reference when compared with cerebellum reference counterparts. Segmentation and PVEC application led to elevated differences in diagnosis-related COMP-SUVR. The most prominent distinctions between mean SUVR in both MCI/HC (+22%) and AD/HC (+70%) were observed when using COMP-SUVR-PVEC$_{BST}$ ($F = 183$, $p < 0.001$) or COMP-SUVR-PVEC$_{WM}$ (MCI/HC: +15%; AD/HC: +52%; $F = 182$, $p < 0.001$).

SUV changes after VOI-based PVEC

Magnitudes of % change after PVEC differed significantly between diagnosis groups for the GM segmented COMP-VOI ($F = 166$; $p < 0.001$), as SUV decreases were observed for HC (−18 ± 12%) and MCI subjects (−10 ± 15%) while SUV of AD subjects increased after PVEC (+7 ± 13%) (Fig. 2). Differences between the diagnosis groups regarding changes after PVEC were also found in the GM segmented CBL-VOI ($F = 5.1$, $p < 0.01$) with decreases in HC (−8 ± 5%), MCI (−7 ± 5%) and AD (−6 ± 5%) subjects. Mean GM volumes for COMP results were 485 ± 37 cm^3 (HC), 480 ± 41 cm^3 (MCI; $p = 0.08$ vs. HC) and 452 ± 37 cm^3 (AD; $p < 0.001$ vs. HC and MCI). Corresponding GM volumes for CBL resulted in 127 ± 9 cm^3 (HC), 128 ± 8 cm^3 (MCI) and 124 ± 8 cm^3 (AD; $p < 0.001$ vs. HC and MCI).

ROC analyses

COMP-SUVR-REF values showed a high correlation ($R = 0.98$; $p < 0.001$) with the corresponding available values obtained from the ADNI database, which were likewise derived from computation of a GM composite target VOI scaled by the whole cerebellum (Johnson et al., 2013; Joshi et al., 2012). Slightly higher values were obtained in our analyses, which have a linear relation defined by $y = 1.07x + 0.03$ with respect to different segmentation methods.

All values for ROC analyses and comparisons are detailed in Table 4. When using COMP-SUVR-REF values, ROC analyses gave an AUC of 0.823 (95%-CI: 0.786–0.856) for the contrast between AD and HC. Respective AUCs for the MCI/HC and the AD/MCI contrasts were 0.622 (95%-CI: 0.587–0.656) and 0.717 (95%-CI: 0.680–0.751), respectively.

When contrasting AD against HC by method, the highest AUC of 0.907 (95%-CI: 0.876–0.938) was found using the white matter VOI as a reference region and VOI-based PVEC performed after GM-segmentation (COMP-SUVR-PVEC$_{WM}$). The ROC comparison of COMP-SUVR-PVEC$_{WM}$ differed very significantly from that of COMP-SUVR-REF ($p < 0.00001$) (Fig. 3A). The AUC for the brainstem reference region was 0.897 (95%-CI: 0.864–0.929) when using VOI-based PVEC after GM segmentation ($p < 0.00001$ vs. COMP-SUVR-REF) (Fig. 3B), but only 0.778 (95%-CI: 0.738–0.815) when using full atlas VOIs without PVEC and the whole cerebellum as a reference region ($p < 0.00001$ vs. COMP-SUVR-REF) (Fig. 3C).

For the contrast of MCI against HC, the AUC for the white matter reference region was 0.651 (95%-CI: 0.613–0.688) when using VOI-based PVEC after GM segmentation ($p =$ n.s. vs. COMP-SUVR-REF) (Fig. 3D). The highest AUC of 0.658 (95%-CI: 0.620–0.696) was obtained through use of brainstem as reference region and performance of VOI-based PVEC after GM-segmentation (COMP-SUVR-PVEC$_{BST}$). The ROC comparison of COMP-SUVR-PVEC$_{BST}$ against COMP-SUVR-REF revealed a significant difference ($p < 0.01$) (Fig. 3E), while the AUC was as low as 0.610 (95%-CI: 0.575–0.644) when using full atlas VOIs without PVEC and the whole cerebellum as a reference region ($p =$ n.s. vs. COMP-SUVR-REF) (Fig. 3F).

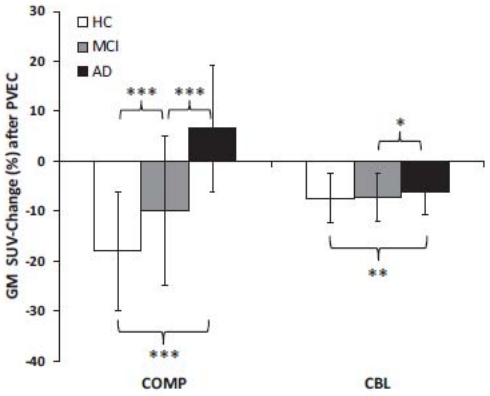


Fig. 2. SUV changes after PVEC. % (± SD) changes after VOI-based partial volume effect correction (PVEC) in the gray matter segmented composite VOI (COMP) and the cerebellar gray matter (CBL). Mean values are given for HC (white; N = 316), MCI (gray; N = 483) and AD (black; N = 163) subjects. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Longitudinal analyses

Findings in stable Aβ-negative subjects

Correlation analyses in 74 stable Aβ-negative subjects revealed lowest agreement between baseline and two year follow-up COMP-SUVR for the cerebellar reference region (FULL: $R = 0.68$; GM: $R = 0.55$; PVEC: $R = 0.60$; Fig. 4A). Scaling to the brainstem uptake resulted at

Table 3
Mean (± SD) COMP-SUVR when assessed by the ten different methodological configurations (column 1) by diagnosis groups (columns 2–4) as well as % differences between diagnosis groups (columns 5–6). F-statistics and p-values for diagnosis related MANOVA are presented in columns 7–8.

Configuration	HC	MCI	AD	MCI vs. HC	AD vs. HC	F-statistic	p-value
COMP-SUVR-FULL$_{CBL}$	1.29 ± 0.17	1.36 ± 0.20	1.48 ± 0.21	5.6%	14.9%	54.4	<0.001
COMP-SUVR-FULL$_{BST}$	0.86 ± 0.11	0.93 ± 0.14	1.05 ± 0.16	7.6%	21.9%	106.6	<0.001
COMP-SUVR-FULL$_{WM}$	0.92 ± 0.04	0.95 ± 0.06	1.01 ± 0.04	3.3%	9.8%	143.6	<0.001
COMP-SUVR-GM$_{CBL}$	1.29 ± 0.20	1.39 ± 0.25	1.55 ± 0.25	7.8%	21.0%	71.3	<0.001
COMP-SUVR-GM$_{BST}$	0.82 ± 0.13	0.90 ± 0.17	1.08 ± 0.18	10.3%	31.9%	146.8	<0.001
COMP-SUVR-GM$_{WM}$	0.88 ± 0.08	0.92 ± 0.10	1.03 ± 0.08	4.5%	17.0%	159.1	<0.001
COMP-SUVR-PVEC$_{CBL}$	1.16 ± 0.35	1.38 ± 0.47	1.79 ± 0.47	19.0%	53.9%	111.0	<0.001
COMP-SUVR-PVEC$_{BST}$	0.69 ± 0.20	0.84 ± 0.28	1.17 ± 0.29	22.2%	70.2%	183.4	<0.001
COMP-SUVR-PVEC$_{WM}$	0.73 ± 0.17	0.85 ± 0.23	1.11 ± 0.21	15.1%	52.1%	181.9	<0.001
COMP-SUVR-REF	1.22 ± 0.19	1.32 ± 0.24	1.52 ± 0.24	8.3%	24.5%	96.0	<0.001

Table 4
ROC calculations for AD/HC (A), MCI/HC (B) and AD/MCI (C) discrimination by the ten different methodological configurations (column 1). Respective areas under the curve (AUC) with standard errors (SE) and 95% confidence interval (CI) are depicted in columns 2–5. All ROCs from our analyses were compared (DeLong et al., 1988) against the widely used reference method COMP-SUVR-REF as (column 6). [^{18}F]-AV45 cut-off values, sensitivities and specificities (best tradeoff) for the discrimination between diagnosis groups are provided in columns 7–9.

Configuration	AUC	SE	Lower 95% CI	Higher 95% CI	p-value	CUT-OFF	Sensitivity	Specificity
A) AD vs HC								
COMP-SUVR-REF	0.823	0.022	0.786	0.856		1.36	80%	80%
COMP-SUVR-FULL$_{CBL}$	0.778	0.024	0.738	0.815	0.00000	1.36	75%	75%
COMP-SUVR-GM$_{CBL}$	0.799	0.023	0.760	0.834	0.00000	1.39	78%	78%
COMP-SUVR-PVEC$_{CBL}$	0.848	0.020	0.813	0.879	0.00065	1.42	82%	82%
COMP-SUVR-FULL$_{BST}$	0.826	0.022	0.789	0.859	0.76820	0.95	81%	81%
COMP-SUVR-GM$_{BST}$	0.860	0.020	0.825	0.890	0.00105	0.94	82%	82%
COMP-SUVR-PVEC$_{BST}$	0.897	0.017	0.864	0.929	0.00000	0.89	85%	83%
COMP-SUVR-FULL$_{WM}$	0.902	0.016	0.870	0.933	0.00000	0.97	85%	85%
COMP-SUVR-GM$_{WM}$	0.907	0.016	0.877	0.938	0.00000	0.96	84%	84%
COMP-SUVR-PVEG$_{WM}$	0.907	0.016	0.876	0.938	0.00000	0.93	85%	85%
B) MCI vs HC								
COMP-SUVR-REF	0.622	0.020	0.587	0.656		1.19	61%	61%
COMP-SUVR-FULL$_{CBL}$	0.610	0.020	0.575	0.644	0.05198	1.27	59%	58%
COMP-SUVR-GM$_{CBL}$	0.621	0.020	0.586	0.654	0.75479	1.25	60%	61%
COMP-SUVR-PVEC$_{CBL}$	0.639	0.020	0.604	0.672	0.04346	1.07	61%	60%
COMP-SUVR-FULL$_{BST}$	0.641	0.019	0.606	0.674	0.15284	0.87	60%	63%
COMP-SUVR-GM$_{BST}$	0.647	0.019	0.613	0.681	0.03741	0.80	64%	60%
COMP-SUVR-PVEC$_{BST}$	0.658	0.019	0.620	0.696	0.00922	0.65	62%	62%
COMP-SUVR-FULL$_{WM}$	0.632	0.019	0.594	0.670	0.53321	0.93	60%	58%
COMP-SUVR-GM$_{WM}$	0.632	0.019	0.594	0.670	0.54871	0.88	60%	58%
COMP-SUVR-PVEG$_{WM}$	0.651	0.019	0.613	0.688	0.07752	0.72	62%	60%
C) AD vs MCI								
COMP-SUVR-REF	0.717	0.025	0.680	0.751		1.45	69%	70%
COMP-SUVR-FULL$_{CBL}$	0.674	0.026	0.636	0.710	0.00000	1.41	65%	64%
COMP-SUVR-GM$_{CBL}$	0.691	0.025	0.653	0.726	0.00000	1.47	66%	66%
COMP-SUVR-PVEC$_{CBL}$	0.731	0.024	0.694	0.764	0.02492	1.61	70%	69%
COMP-SUVR-FULL$_{BST}$	0.724	0.025	0.688	0.758	0.60534	1.01	68%	68%
COMP-SUVR-GM$_{BST}$	0.760	0.024	0.725	0.793	0.00025	1.02	72%	72%
COMP-SUVR-PVEC$_{BST}$	0.792	0.021	0.751	0.833	0.00000	0.92	74%	74%
COMP-SUVR-FULL$_{WM}$	0.792	0.020	0.753	0.830	0.00000	1.00	74%	74%
COMP-SUVR-GM$_{WM}$	0.805	0.019	0.767	0.843	0.00000	1.00	74%	73%
COMP-SUVR-PVEG$_{WM}$	0.803	0.020	0.764	0.842	0.00000	1.03	74%	74%

an intermediate correlation (FULL: R = 0.82; GM: R = 0.80; PVEC: R = 0.81; Fig. 4B), while the WM reference showed the highest agreement (FULL: R = 0.92; GM: R = 0.93; PVEC: R = 0.87; Fig. 4C). The COMP-SUVR-REF configuration showed an R of 0.58.

Changes in GM volume and longitudinal [^{18}F]-AV45 uptake

To two-year follow-up, GM Δ%-volumes for COMP decreased −1.0 ± 4.3% in HC, −1.9 ± 3.0% in MCI and −3.3 ± 4.4% in AD ($p < 0.01$ vs. HC). Corresponding GM Δ%-volumes for CBL decreased −1.3 ± 2.8% in HC, −1.4 ± 2.2% in MCI and −1.8 ± 3.2% in AD (all n.s.). Low longitudinal changes over two years were found in reference region SUVs of CBL-FULL-VOI (+0.1 −0.2%), CBL-GM-VOI (+0.3–0.6%), BST-VOI (+0.2–0.7%), and WM-VOI (+0.1–0.2%), with no differences between diagnosis groups. Longitudinal SUV decreases were observed in a small subset of nine converting HC (CBL-FULL: −2.0 ± 2.6%; CBL-GM: −1.2 ± 1.8%; BST: −4.3 ± 6.2%; WM: −4.7 ± 4.4%) and twelve converting MCI (CBL-FULL: −1.3 ± 2.8%; CBL-GM: −0.1 ± 2.2%; BST: −2.4 ± 6.3%; WM: −2.5 ± 8.1%) subjects. All mean [^{18}F]-AV45 changes (% ± SD) to follow-up for the ten differently-assessed Δ-COMP-SUVRs are presented in Table 5. Greatest increases over time were observed when using PVEC regardless of the reference region (Fig. 5A). When only Aβ-positive subjects were considered, highest increases were again found when using PVEC (Fig. 5B). Inter-subject variability was consequently lowest for white matter, intermediate for brainstem and highest for cerebellum as reference region (Fig. 5A + B).

Discussion

We present a systematic investigation of the discriminatory performance of different reference regions for [^{18}F]-AV45-PET quantitation, making use of the hitherto largest dataset of PET examinations in AD patients, along with MCI and HC groups. In addition, we tested effects of additionally performed MRI-based segmentation and VOI-based PVEC in our nearly 1000 cases. ROC analyses revealed that the discriminatory power between MCI or AD and the healthy control population was increased by using the white matter or the brainstem rather than cerebellum as reference region, and was also increased by application of a VOI-based PVEC. However, for the particular case of longitudinal analyses, white matter emerged as the best-suited reference region, due to the relatively low group variability it provided. Furthermore, the expected increases with disease progression in cortical amyloidosis were best captured using additional GM segmentation and PVEC.

Reference region

White matter has recently been evaluated as a reference region for [^{18}F]-FDDNP PET in a small sample of AD patients (Wong et al., 2010). The authors of that study concluded that more accurate and less biased binding estimates could be established when using a white matter reference instead of the cerebellum. In our study, cerebellum and brainstem reference quantitation showed significantly lower SUV in AD subjects when compared to HC. In this contrast, which is critical for monitoring early disease progression, the white matter reference gave the more stable SUV results between diagnosis groups at baseline, and therefore performed best with regard to invariance with the investigated pathology, i.e. GM amyloidosis.

The lower [^{18}F]-AV45 SUVs observed in cerebellum and brainstem were not expected in AD subjects, based on knowledge of the distribution of amyloid, and thus need to be further investigated. Interestingly, we found decreases in all reference region SUVs in the group of subjects

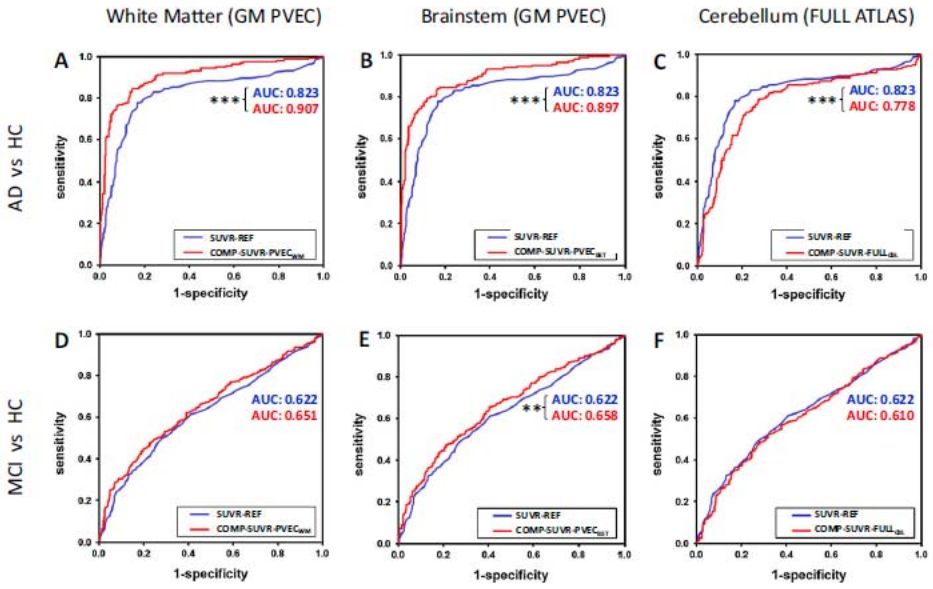


Fig. 3. ROC analyses for AD/HC and MCI/HC discrimination. ROCs were calculated and compared for discriminations by [^{18}F]-AV45 uptake between AD and HC (upper row, A–C) as well as between MCI and HC (lower row, D–F) subjects. Exemplary ROCs for COMP-SUVR-PVEC$_{WM}$ (column 1), COMP-SUVR-PVEC$_{BST}$ (column 2) and COMP-SUVR-FULL$_{CBL}$ (column 3) configurations (all red) were consequently contrasted against the COMP-SUVR-REF configuration (blue). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

converting from HC to MCI ($N = 9$) or MCI to AD ($N = 12$). However these data are not in accord with stable results between diagnosis groups in the WM SUV, as this region was affected as well, a phenomenon which is probably related to the small sample size of converters. A higher inter-subject variation (SD: 15.1%) in WM SUV was observed when compared to brainstem SUV (SD: 12.1%) or cerebellum SUV (SD: 5.8%) regardless of the diagnosis group. From this side most stable cortical quantification would have been expected from the cerebellar reference. However, the longitudinal correlation of COMP-SUVR between baseline and 2-year follow-up in 74 Aβ-negative subjects was by far the highest when using the WM reference region (Fig. 4). These findings in presumably pathology free patients support the suitability of WM reference for a stable cortical quantification within individual subjects.

Interestingly, the WM reference region including not only true WM voxels performed better than did the WM-seg reference region. For example, the CSF space in the depths of sulci was present in the WM VOI configuration, arising from the volumes in the full atlas. Therefore, some extra-cerebral volumes contributed to those "non-true" WM voxels. The longitudinal investigation of amyloid negative subjects resulted in a R of 0.85 when correlating baseline versus 2-year follow-up of COMP-GM-SUV/"non-true" WM. We note that using these "non-true" WM voxels as a stand-alone reference region did not result in a performance level quite as high as for WM (which had an $R = 0.93$), but that the correlation was still higher than we had expected. We suppose that WM reference region may be stabilized by inclusion of the "non-true" WM voxels lying adjacent to the GM target region. In particular, spill-over from variable amounts of nonspecific GM uptake may be counterbalanced by including voxels adjacent to the GM (e.g. outer CSF space) in the reference region. Alternately, we suppose that inter-subject variability in nonspecific [^{18}F]-AV45 binding in "non-true" WM voxels would influence the signal attributed to GM.

A recent methodologically-oriented amyloid PET study comparing [^{18}F]-AV45, [^{18}F]-flutemetamol and [^{11}C]-PiB has also calculated composite SUVR with a brainstem reference (Landau et al., 2014). The [^{18}F]-AV45 findings in their 32 cases showed consequently lower composite SUVR$_{BST}$ compared to SUVR$_{CBL}$, which is in line with our results, as were also their estimates for [^{18}F]-AV45 in mostly MCI subjects for the magnitudes of SUVR$_{BST}$ (0.88 vs. 0.90) and SUVR$_{CBL}$ (1.38 vs. 1.39). In another methods study, Edison showed that a target-to-pons ratio for [^{11}C]-PiB offered reliable estimations of the cortical radiotracer uptake, especially in cases where use of the cerebellum is not appropriate, such as in familial AD (Edison et al., 2012). Although the rather low target-to-WM or target-to-BST ratios, which were frequently < 1, may perplex the observer, we can only agree with these earlier findings, as we likewise obtained very stable and robust results for all assessed COMP-SUVR$_{WM}$ or COMP-SUVR$_{BST}$, which gave an equal relative standard deviation when compared to COMP-SUVR$_{CBL}$.

Our ROC analyses proved better discrimination between the present large groups of AD, MCI and HC subjects using the white matter or the brainstem as a reference region (Fig. 3). Indeed, the differentiation of these clinical groups is an important precondition for preventive therapy trials. Another important observation of our study was the notably lower AUC ($p < 0.001$ compared to COMP-SUVR-REF) when plain atlas VOIs without GM segmentation or PVEC were used, a finding which was also found for whole cerebellum as a reference tissue. Use of precisely this approach therefore probably lowers the discriminatory power between MCI/HC and AD/HC, when compared to the analysis strategies investigated for the initial validation of this tracer (Clark et al., 2012; Johnson et al., 2013). The use of a brainstem reference with full atlas VOIs showed comparable results with COMP-SUVR-REF, and can therefore serve adequately even when segmentation/PVEC are not available (e.g. without MRI information).

VOI-based PVEC

In an extensive review, Erlandsson emphasized the effects of marked cortical atrophy in neurodegenerative diseases, which lead to exaggerated PVE when performing brain PET studies (Erlandsson

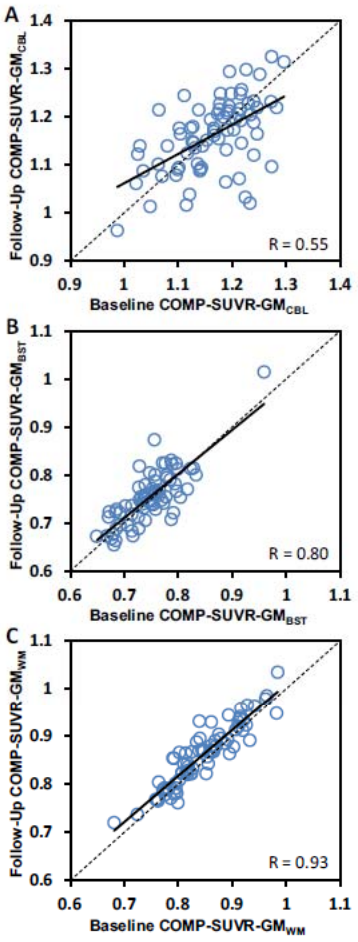


Fig. 4. Longitudinal findings in stable Aβ-negative subjects. Agreement between baseline and follow-up COMP-SUVR-GM for 74 (N = 33 HC, N = 39 MCI, N = 2 AD) stable non-amyloid accumulating subjects. Scaling was performed either by the cerebellar gray matter (A), the brainstem (B) or the white matter (C). Trend lines of correlation are illustrated continuously while dotted lines represent assumed perfect identity. R give estimated Pearson's coefficients of correlation.

et al., 2012). Even apparent hypometabolism in normal aging as detected by [^{18}F]-FDG PET was partially recoverable by MRI-based PVEC (Curiati et al., 2011), and is therefore likely in large part a bias due to atrophy and loss of signal. On the other hand, age-related increases in oxygen extraction fraction are likely robust to atrophy and PVE (Aanerud et al., 2012).

For [^{18}F]-flutemetamol amyloid PET data, a region-based voxel-wise PVEC proved suitable (Thomas et al., 2011) after validation by phantom experiments. The authors noted that not only atrophy and the spill-over from a region, but also the spill-in must be considered when performing PVEC. They concluded that techniques accounting for multiple regions and tissue types are more appropriate for amyloid PET, especially given that significant WM uptake is a well-known attribute of [^{18}F]-flutemetamol, [^{18}F]-AV45 and [^{18}F]-florbetaben. We concur with their position, and recently undertook a microPET study testing effects of VOI-based PVEC of [^{18}F]-florbetaben uptake in brain of AD transgenic mouse brain (Brendel et al., 2014). We were able to verify by gamma counting and autoradiography ex vivo (where PVE is avoided) that distinct overcorrections occur when VOI configurations for the mouse brain lack WM and background regions; these regions were consequently included in the VOI-based PVEC approach of human brain in the current work, which yielded plausible values and changes after PVEC.

Present results indicate decreased mean changes of cortical SUV in amyloid tracer measurements after PVEC in HC and MCI subjects, whereas AD subjects showed an increase, no doubt due to their greater atrophy. Concerning the high white matter uptake of [^{18}F]-AV45, which is present irrespective of brain amyloidosis, this observation also makes sense; in most HC cases, the low cortical tracer uptake is located adjacent to the distinctly higher tracer uptake in underlying WM, which leads to net spill-in to GM in cases of low-moderate atrophy. A VOI-based PVEC approach including a white matter VOI accommodates this, and consequently leads to a decrease after PVEC in HCs. On the other hand, AD subjects show cortical [^{18}F]-AV45 uptake equal to or exceeding uptake in the subjacent WM, in conjunction with severe cortical atrophy. Therefore recovery of [^{18}F]-AV45 signal in GM necessarily increases after PVEC. MCI subjects fall somewhere between these two extremes, and in the present investigation showed as a group a small net decrease after PVEC, which is probably related to a quite high proportion of Aβ-negative cases (45%) in MCI groups.

Impact on longitudinal approaches

Longitudinal amyloid imaging recently showed high inter-subject variability when methodologically investigated with [^{11}C]-PiB PET to mean 2.5-year follow-up, and some AD subjects even showed an apparent decrease in amyloid burden over time when assessed by SUVR methods, which seems at odds with the known pathophysiology of disease progression (van Berckel et al., 2013). However, this seeming decrease was not apparent when using the full dynamic data set for calculation of binding potentials (BP_{ND}), suggesting blood flow changes between scans. We likewise observed high relative standard deviations in Δ%-COMP-SUVR for [^{18}F]-AV45 with two years of progression in our cohort. Two relevant effects were observed in our methodological comparison: First, the use of individual segmented GM VOIs (scaled by CBL and BST) showed higher Δ%-COMP-SUVR compared to full atlas VOIs, while additional performance of the PVEC substantially increased this effect, leading to a mean Δ%-COMP-SUVR of approximately 5.5% over two years for amyloid accumulation in HC and MCI subjects. Second, the use of the white matter reference distinctly lowered the relative standard deviations of diagnosis-related Δ%-COMP-SUVR group means, thus favoring the discrimination by group. The results of the hitherto largest longitudinal [^{11}C]-PiB PET study indicated an increase of 1.5% per year in HC and MCI subjects with low [^{11}C]-PiB levels and 3.1% per year in HC and MCI subjects with high [^{11}C]-PiB retention with cerebellar reference and without use of PVEC (Villemagne et al., 2013). In our analysis, the maximum mean [^{18}F]-AV45 increase observed in configurations without PVEC and with a cerebellar reference did not exceed 1.5% per year for amyloid accumulation in HC and MCI subjects. Taking into consideration the comparably high WM uptake of [^{18}F]-AV45, we can speculate that the detection of real longitudinal differences is suppressed by both spillover and progressing atrophy when PVEC is not performed. Thus, our findings lead us to conjecture that the fitness of quantitative longitudinal [^{18}F]-AV45 assessment over a brief two year period mainly benefits from the use of PVEC and secondarily from stabilization by the WM reference, which propagates to reduced inter-subject variability, i.e. improved precision. Although doubtless significant in planned intervention studies, we are aware that such differences are of minor relevance to routine clinical decisions with rating of Aβ-positivity/Aβ-negativity.

Table 5
Mean (± SD) % changes over two years in [^{18}F]-AV45 uptake when assessed by the ten different COMP-SUVR methodological configurations (column 1) for the three diagnosis groups (columns 2–4) for all subjects. Columns 5–7 display corresponding results when only Aβ-positive subjects were considered. Significant differences of all different configurations when contrasted against the COMP-SUVR-REF configuration are indicated by: *p < 0.05; **p < 0.01; ***p < 0.001.

Configuration	All subjects			Aβ-positive subjects		
	HC (N = 88)	MCI (N = 148)	AD (N = 22)	HC (N = 26)	MCI (N = 66)	AD (N = 16)
2-year-Δ-COMP-SUVR-FULL$_{CBL}$	1.0 ± 5.5%***	0.6 ± 6.3%***	−0.4 ± 8.8%**	2.1 ± 7.2%***	0.9 ± 7.7%***	−1.2 ± 7.7%***
2-year-Δ-COMP-SUVR-FULL$_{BST}$	1.0 ± 4.3%	0.5 ± 4.5%*	−1.5 ± 4.3%	2.4 ± 5.0%	1.2 ± 4.8%	−1.1 ± 4.7%
2-year-Δ-COMP-SUVR-FULL$_{WM}$	0.9 ± 1.6%	1.2 ± 1.7%	1.1 ± 1.3%	1.0 ± 1.2%	1.6 ± 1.6%	1.4 ± 1.1%
2-year-Δ-COMP-SUVR-GM$_{CBL}$	1.6 ± 6.3%	1.3 ± 7.0%***	0.6 ± 9.6%	2.9 ± 8.3%	1.8 ± 8.5%***	−0.6 ± 8.5%*
2-year-Δ-COMP-SUVR-GM$_{BST}$	1.9 ± 4.5%	1.7 ± 4.7%	−0.1 ± 4.3%	3.5 ± 5.3%	2.9 ± 4.9%	0.6 ± 4.6%
2-year-Δ-COMP-SUVR-GM$_{WM}$	1.7 ± 3.0%	2.4 ± 3.4%	2.4 ± 2.9%	2.1 ± 2.5%	3.3 ± 3.5%	3.0 ± 2.4%
2-year-Δ-COMP-SUVR-PVEC$_{CBL}$	4.6 ± 9.9%***	4.2 ± 10.6%***	4.2 ± 12.7%*	6.5 ± 12.7%*	5.4 ± 12.2%***	2.7 ± 12.7%
2-year-Δ-COMP-SUVR-PVEC$_{BST}$	3.6 ± 7.5%	3.3 ± 7.1%*	1.6 ± 6.3%	5.6 ± 6.5%	5.2 ± 7.1%**	2.8 ± 6.0%
2-year-Δ-COMP-SUVR-PVEC$_{WM}$	3.4 ± 6.5%	4.0 ± 6.7%***	4.2 ± 5.9%	4.2 ± 4.2%	5.5 ± 6.3%*	5.1 ± 4.2%*
2-year-Δ-COMP-SUVR-REF	1.9 ± 5.4%	1.8 ± 6.3%	1.0 ± 8.2%	3.2 ± 7.5%	2.6 ± 7.8%	0.5 ± 7.7%

Limitations

Data included in this study derived from multiple centers. Therefore different quality of PET and MRI data was obtained and could only partially be balanced by image preprocessing. However it has to be considered that the same study conditions will likely occur in multicenter treatment trials. Furthermore, the 17 different PET scanners were included as a covariate to adjust for the possibility of instrumentation bias, which resulted in only small changes (<1%) between adjusted and non-adjusted data.

The detection of longitudinal reductions of amyloid burden, as seen with the SUVR method in some configurations, might suffer from interference from inadequate quantitation, since blood-flow changes between scans might not be properly accommodated. However, in large-scale studies, the use of SUVR is a convenient proxy for quantification of brain amyloid burden. Ultimately, the magnitude of this bias could only be answered by direct assessment of dynamic datasets in patients studied longitudinally, or through concomitant quantitative CBF studies. It seems premature to condemn SUVRs entirely, insofar as SUVRs using the three [^{18}F]-amyloid tracers have proven to correlate well with histopathological comparisons, as assessed in the development of these analyses.

Diagnosis groups of HC, MCI and AD were obtained by clinical assessment, with no histopathological gold standard information available. Therefore discriminatory analyses derive from detailed neuropsychiatric test batteries and biomarkers, with as yet no final validation. Furthermore, the detected longitudinal increases of amyloid burden over time could only be compared with previous findings and extrapolations of a large prospective PET study (Villemagne et al., 2013), since histopathologically verified amyloid accumulation over time is still lacking.

As not all extracortical regions in proximity to the brain (e.g. fat, bone, scalp) were included in the VOI-based algorithm used for PVEC, spillover from these sources may possibly contribute to the attribution of signal from brain VOIs; however, this potential bias seems likely to be of low magnitude since the distance from extracerebral structures is in the order of the FWHM. Our findings are for [^{18}F]-AV45, and may not generalize for other amyloid radioligands, which will have to be investigated in this context in further studies.

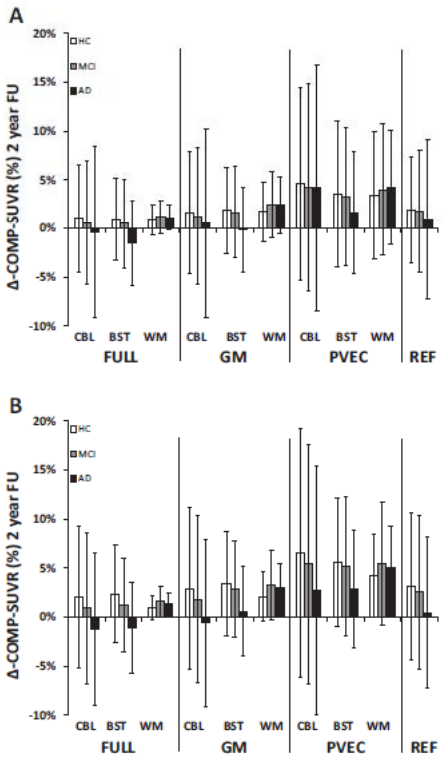


Fig. 5. Changes in longitudinal [^{18}F]-AV45 uptake. Longitudinal % changes in [^{18}F]-AV45 uptake (± SD) as assessed for HC (white; N = 88), MCI (gray; N = 148) and AD (black; N = 22) subjects and calculated for the ten COMP-SUVR configurations (A), as well as corresponding longitudinal % changes in [^{18}F]-AV45 uptake (± SD) when only Aß positive subjects were considered (B). Either full atlas VOIs (FULL), gray matter segmented VOIs (GM) or gray matter segmented VOIs after partial volume effect correction (PVEC) were used and scaled by either the cerebellum (CBL), the brainstem (BST) or the white matter (WM). An additional standard configuration consisted of GM segmented VOIs scaled by the whole cerebellum (REF).

Conclusion

We aimed to optimize the discriminative fitness of [^{18}F]-AV45-PET by methodological considerations of cortical atrophy/spill-over and comparison of reference regions. We found clear benefits from using VOI-based PVEC and defining the white matter or the brainstem rather than cerebellum as the reference region; both procedures increased the discriminatory power between MCI and HC, which are key target groups in therapy trials. Furthermore, longitudinal increases in amyloidosis during follow-up for two years were likely more accurately captured by application of the present optimized procedure with the white matter reference.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.11.055.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to Rev December 5, 2013 support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The authors acknowledge Inglewood Biomedical Editing for professional editing of the manuscript.

References

Aanerud, J., Borghammer, P., Chakravarty, M.M., Vang, K., Rodell, A.B., Jonsdottir, K.Y., Moller, A., Ashkanian, M., Vafaee, M.S., Iversen, P., Johannsen, P., Gjedde, A., 2012. Brain energy metabolism and blood flow differences in healthy aging. J. Cereb. Blood Flow Metab. 32, 1177–1187.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bohnen, N.I., Djang, D.S., Herholz, K., Anzai, Y., Minoshima, S., 2012. Effectiveness and safety of ^{18}F-FDG PET in the evaluation of dementia: a review of the recent literature. J. Nucl. Med. 53, 59–71.

Borghammer, P., Cumming, P., Aanerud, J., Forster, S., Gjedde, A., 2009. Subcortical elevation of metabolism in Parkinson's disease--a critical reappraisal in the context of global mean normalization. Neuroimage 47, 1514–1521.

Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Brendel, M., Delker, A., Rotzer, C., Boning, G., Carlsen, J., Cyran, C., Mille, E., Gildehaus, F.J., Cumming, P., Baumann, K., Steiner, H., Haass, C., Herms, J., Bartenstein, P., Rominger, A., 2014. Impact of partial volume effect correction on cerebral beta-amyloid imaging in APP-Swe mice using [(18)F]-florbetaben PET. Neuroimage 84, 843–853.

Clark, C.M., Pontecorvo, M.J., Beach, T.G., Bedell, B.J., Coleman, R.E., Doraiswamy, P.M., Fleisher, A.S., Reiman, E.M., Sabbagh, M.N., Sadowsky, C.H., Schneider, J.A., Arora, A., Carpenter, A.P., Flitter, M.L., Joshi, A.D., Krautkramer, M.J., Lu, M., Mintun, M.A., Skovronsky, D.M., 2012. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. Lancet Neurol. 11, 669–678.

Curiati, P.K., Tamashiro-Duran, J.H., Duran, F.L., Buchpiguel, C.A., Squarzoni, P., Romano, D.C., Vallada, H., Menezes, P.R., Scazufca, M., Busatto, G.F., Alves, T.C., 2011. Age-related metabolic profiles in cognitively healthy elders: results from a voxel-based [^{18}F]fluorodeoxyglucose-positron-emission tomography study with partial volume effects correction. AJNR Am. J. Neuroradiol. 32, 560–565.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

Dukart, J., Perneczky, R., Forster, S., Barthel, H., Diehl-Schmid, J., Draganski, B., Obrig, H., Santarnecchi, E., Drzezga, A., Fellgiebel, A., Frackowiak, R., Kurz, A., Muller, K., Sabri, O., Schroeter, M.L., Yakushev, I., 2013. Reference cluster normalization improves detection of frontotemporal lobar degeneration by means of FDG-PET. PLoS One 8, e55415.

Edison, P., Hinz, R., Ramlackhansingh, A., Thomas, J., Gelosa, G., Archer, H.A., Turkheimer, F.E., Brooks, D.J., 2012. Can target-to-pons ratio be used as a reliable method for the analysis of [11C]PIB brain scans? Neuroimage 60, 1716–1723.

Erlandsson, K., Buvat, I., Pretorius, P.H., Thomas, B.A., Hutton, B.F., 2012. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. Phys. Med. Biol. 57, R119–R159.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–247.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Johnson, K.A., Sperling, R.A., Gidicsin, C.M., Carmasin, J.S., Maye, J.E., Coleman, R.E., Reiman, E.M., Sabbagh, M.N., Sadowsky, C.H., Fleisher, A.S., Murali Doraiswamy, P., Carpenter, A.P., Clark, C.M., Joshi, A.D., Lu, M., Grundman, M., Mintun, M.A., Pontecorvo, M.J., Skovronsky, D.M., group, A.A.s, 2013. Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. Alzheimers Dement. 9, S72–S83.

Joshi, A.D., Pontecorvo, M.J., Clark, C.M., Carpenter, A.P., Jennings, D.L., Sadowsky, C.H., Adler, L.P., Kovnat, K.D., Seibyl, J.P., Arora, A., Saha, K., Burns, J.D., Lowrey, M.J., Mintun, M.A., Skovronsky, D.M., Florbetapir, F.S.I., 2012. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. J. Nucl. Med. 53, 378–384.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.

Landau, S.M., Thomas, B.A., Thurfjell, L., Schmidt, M., Margolin, R., Mintun, M., Pontecorvo, M., Baker, S.L., Jagust, W.J., 2014. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur. J. Nucl. Med. Mol. Imaging 41, 1398–1407.

Lim, Y.Y., Ellis, K.A., Pietrzak, R.H., Ames, D., Darby, D., Harrington, K., Martins, R.N., Masters, C.L., Rowe, C., Savage, G., Szoeke, C., Villemagne, V.L., Maruff, P., Group, A.R., 2012. Stronger effect of amyloid load than APOE genotype on cognitive decline in healthy older adults. Neurology 79, 1645–1652.

Lim, Y.Y., Maruff, P., Pietrzak, R.H., Ellis, K.A., Darby, D., Ames, D., Harrington, K., Martins, R.N., Masters, C.L., Szoeke, C., Savage, G., Villemagne, V.L., Rowe, C.C., Group, A.R., 2014. Abeta and cognitive change: examining the preclinical and prodromal stages of Alzheimer's disease. Alzheimers Dement. 10, 743–751.

Rousset, O.G., Ma, Y., Evans, A.C., 1998. Correction for partial volume effects in PET: principle and validation. J. Nucl. Med. 39, 904–911.

Rousset, O.G., Collins, D.L., Rahmim, A., Wong, D.F., 2008. Design and implementation of an automated partial volume correction in PET: application to dopamine receptor quantification in the normal human striatum. J. Nucl. Med. 49, 1097–1106.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87–97.

Thomas, B.A., Erlandsson, K., Modat, M., Thurfjell, L., Vandenberghe, R., Ourselin, S., Hutton, B.F., 2011. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. Eur. J. Nucl. Med. Mol. Imaging 38, 1104–1119.

van Berckel, B.N., Ossenkoppele, R., Tolboom, N., Yaqub, M., Foster-Dingley, J.C., Windhorst, A.D., Scheltens, P., Lammertsma, A.A., Boellaard, R., 2013. Longitudinal amyloid imaging using ^{11}C-PiB: methodologic considerations. J. Nucl. Med. 54, 1570–1576.

Villemagne, V.L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K.A., Salvado, O., Szoeke, C., Macaulay, S.L., Martins, R., Maruff, P., Ames, D., Rowe, C.C., Masters, C.L., 2013. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 12, 357–367.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Shen, L., Siuciak, J.A., Soares, H., Toga, A.W., Trojanowski, J.Q., 2013. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement. 9, e111–e194.

Wong, K.P., Wardak, M., Shao, W., Dahlbom, M., Kepe, V., Liu, J., Satyamurthy, N., Small, G.W., Barrio, J.R., Huang, S.C., 2010. Quantitative analysis of [^{18}F]FDDNP PET using subcortical white matter as reference region. Eur. J. Nucl. Med. Mol. Imaging 37, 575–588.

Yakushev, I., Landvogt, C., Buchholz, H.G., Fellgiebel, A., Hammers, A., Scheurich, A., Schmidtmann, I., Gerhard, A., Schreckenberger, M., Bartenstein, P., 2008. Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. Psychiatry Res. 164, 143–153.

Ziegler-Graham, K., Brookmeyer, R., Johnson, E., Arrighi, H.M., 2008. Worldwide variation in the doubling time of Alzheimer's disease incidence rates. Alzheimers Dement. 4, 316–323.

Supplement 1

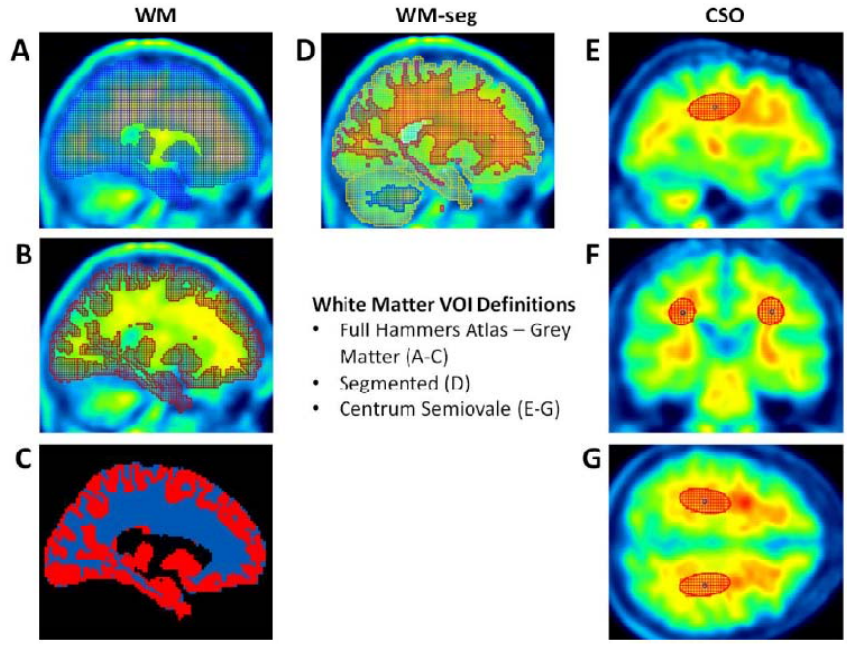
PET image pre-processing

Raw PET images from all sites are downloaded for quality control at University of Michigan. There, raw images are converted to a standard file format. Separate frames are extracted from the image file for registration purposes. Four five-minute frames (ADNI GO/2) are acquired at 50 to 70 minutes post-injection. Each extracted frame is co-registered to the first extracted frame of the raw image file (frame acquired at 50-55 min post-injection). The base frame image and the five co-registered frames (or all co-registered frames for the quantitative studies) are recombined into a co-registered dynamic image set. These image sets have the same image size (for example, 128×128×63) and voxel dimensions (for example, 2.0×2.0×2.0 mm), and remain in the same spatial orientation as the original PET image data. This next step of processing is performed simply by averaging the four five-minute frames of the "Co-registered Dynamic" image set. This creates a single 20 min PET image, still positioned in "native" space. Each subject's co-registered, averaged image from their baseline PET scan is then reoriented into a standard 160×160×96 voxel image grid, having 1.5 mm cubic voxels. This image grid is oriented such that the anterior-posterior axis of the subject is parallel to the AC-PC line. This is referred to as "AC-PC" space in the LONI search program. This standardized image then serves as a reference image for all PET scans on that subject. The individual frames from each PET scan (the baseline study, as well as all subsequent studies (6-month scan, 12-month scan, etc.) are co-registered to this baseline reference image. By doing the co-registration from the original raw image data to a standardized space in a single step, only one interpolation of the image data is required, and thus resolution degradation by interpolation is kept to a

minimum, and is the same for all scans. An averaged image is generated from the "AC-PC" co-registered frames and then intensity-normalized using a subject-specific mask so that the average intensity of voxels within the mask is exactly one. Both the spatial orientation (AC-PC) and the intensity normalization of the image are intended as a starting point for subsequent analyses. With a standardized image matrix, PET data from different scanner models can be compared more easily. It should be noted that in these image sets only spatial re-orientation and intensity normalization of scans has occurred. No non-linear warping or even linear scaling of the brain dimensions has been applied to the images. Uniform resolution is guaranteed by smoothing of the above-mentioned images. Each image set is filtered with a scanner-specific filter function (can be a non-isotropic filter) to produce images of a uniform isotropic resolution of 8 mm FWHM, the approximate resolution of the lowest resolution scanners of the 17 instruments used in ADNI. Image sets from higher resolution scanners obviously will have been smoothed more than those from lower resolution scanners. The specific filter functions were determined from the Hoffman phantom PET scans that were acquired during the certification process.

Supplement 2

Figure S1: Definition of the three differently assessed white matter VOIs



WM (A-C): Sagittal slices: From the cortical full atlas VOI indicated in blue (A), the area of individual cortical GM VOIs in red (B) was subtracted, thus comprising all subcortical white matter voxels displayed in blue (C), as included in the full atlas VOI. Furthermore, voxels in the ventricles and the cerebellum were also excluded from the resulting individual subcortical WM region.

WM-seg (D): The segmentation process was used to generate WM VOIs containing all voxels declared as WM in the segmentation process. This was performed separately for cortex, displayed in red, and the cerebellum, displayed in blue. SUVR were calculated through division by subcortical WM without the cerebellar WM.

CSO (E-F): Sagittal (E), coronal (F) and axial (G) slices illustrate how the manually drawn bilateral centrum semiovale VOI (6.2 cm^3 each) was defined.

Table S1

SUV Region	HC	MCI	AD	F-statistic	p-value
Centrum Semiovale	2.31 ± 0.34	2.22 ± 0.33	2.07 ± 0.32	26.4	< 0.001
cortical WM (segmented)	1.94 ± 0.28	1.94 ± 0.26	1.94 ± 0.27	0.1	n.s.
cerebellar WM (segmented)	1.78 ± 0.21	1.74 ± 0.20	1.65 ± 0.18	22.8	< 0.001

Mean (± SD) SUV for reference regions (column 1) by diagnosis groups (columns 2 - 4). F-statistics and p-values for diagnosis related MANOVA are presented in columns 5 - 6. Significant differences against HC are indicated by: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table S2

Configuration	HC	MCI	AD	MCI vs. HC	AD vs. HC	F-statistic	p-value
COMP-SUVR-FULL$_{CSO}$	0.61 ± 0.09	0.67 ± 0.09	0.75 ± 0.09	8.7%	22.5%	128.6	< 0.001
COMP-SUVR-GM$_{CSO}$	0.58 ± 0.11	0.65 ± 0.10	0.77 ± 0.09	11.4%	32.7%	163.1	< 0.001
COMP-SUVR-PVEC$_{CSO}$	0.49 ± 0.19	0.60 ± 0.19	0.83 ± 0.18	23.3%	71.0%	187.5	< 0.001
COMP-SUVR-FULL$_{WM-seg}$	0.72 ± 0.05	0.76 ± 0.06	0.79 ± 0.06	4.3%	9.7%	77.2	< 0.001
COMP-SUVR-GM$_{WM-seg}$	0.69 ± 0.07	0.73 ± 0.09	0.82 ± 0.08	6.9%	19.0%	143.6	< 0.001
COMP-SUVR-PVEC$_{WM-seg}$	0.57 ± 0.14	0.67 ± 0.18	0.87 ± 0.17	17.7%	53.3%	179.0	< 0.001

Mean (± SD) COMP-SUVR when assessed by the different methodological configurations (column 1) by diagnosis groups (columns 2 - 4) as well as % differences between diagnosis groups (columns 5 - 6). F-statistics and p-values for diagnosis related MANOVA are presented in columns 7 - 8.

Table S3

A) AD vs HC								
Configuration	AUC	SE	Low-95%-CI	HI-95%-CI	p-value	CUT-OFF	sens	spec
COMP-SUVR-REF	0.823	0.022	0.786	0.856		1.36	80%	80%
COMP-SUVR-FULL$_{CSO}$	0.853	0.021	0.811	0.895	0.04314	0.68	82%	82%
COMP-SUVR-GM$_{CSO}$	0.880	0.019	0.843	0.918	0.00030	0.67	83%	83%
COMP-SUVR-PVEC$_{CSO}$	0.898	0.017	0.864	0.931	0.00000	0.66	85%	84%
COMP-SUVR-FULL$_{WM\text{-}seg}$	0.813	0.023	0.768	0.858	0.49419	0.75	79%	80%
COMP-SUVR-GM$_{WM\text{-}seg}$	0.871	0.020	0.831	0.911	0.00298	0.76	83%	82%
COMP-SUVR-PVEC$_{WM\text{-}seg}$	0.898	0.017	0.865	0.932	0.00000	0.74	84%	84%

B) MCI vs HC								
Configuration	AUC	SE	Low-95%-CI	HI-95%-CI	p-value	CUT-OFF	sens	spec
COMP-SUVR-REF	0.622	0.020	0.587	0.656		1.19	61%	61%
COMP-SUVR-FULL$_{CSO}$	0.662	0.019	0.624	0.699	0.02111	0.61	61%	61%
COMP-SUVR-GM$_{CSO}$	0.663	0.019	0.625	0.700	0.01377	0.57	63%	61%
COMP-SUVR-PVEC$_{CSO}$	0.664	0.019	0.627	0.702	0.00871	0.47	62%	61%
COMP-SUVR-FULL$_{WM\text{-}seg}$	0.644	0.019	0.606	0.682	0.24582	0.73	61%	60%
COMP-SUVR-GM$_{WM\text{-}seg}$	0.658	0.019	0.620	0.695	0.03394	0.68	62%	62%
COMP-SUVR-PVEC$_{WM\text{-}seg}$	0.663	0.019	0.626	0.700	0.01339	0.56	62%	62%

C) AD vs MCI								
Configuration	AUC	SE	Low-95%-CI	HI-95%-CI	p-value	CUT-OFF	sens	spec
COMP-SUVR-REF	0.717	0.025	0.680	0.751		1.45	69%	70%
COMP-SUVR-FULL$_{CSO}$	0.736	0.024	0.690	0.782	0.22878	0.73	72%	71%
COMP-SUVR-GM$_{CSO}$	0.771	0.022	0.727	0.814	0.00037	0.74	74%	73%
COMP-SUVR-PVEC$_{CSO}$	0.790	0.021	0.749	0.832	0.00000	0.75	74%	73%
COMP-SUVR-FULL$_{WM\text{-}seg}$	0.684	0.024	0.637	0.732	0.06879	0.78	68%	66%
COMP-SUVR-GM$_{WM\text{-}seg}$	0.761	0.023	0.717	0.806	0.00398	0.80	74%	74%
COMP-SUVR-PVEC$_{WM\text{-}seg}$	0.792	0.021	0.751	0.833	0.00000	0.82	74%	74%

ROC calculations for AD/HC (A), MCI/HC (B) and AD/MCI (C) discrimination by the different methodological configurations (column 1). Respective areas under the curve (AUC) with standard errors (SE) and 95% confidence intervals (CI) are depicted in columns 2 - 5. All ROCs from our analyses were compared (DeLong et al., 1988)

against the widely used reference method COMP-SUVR-REF, with the p-value displayed in column 6. [^{18}F]-AV45 cut-off values, sensitivities and specificities (best tradeoff) for the discrimination between diagnosis groups are provided in columns 7 - 9.

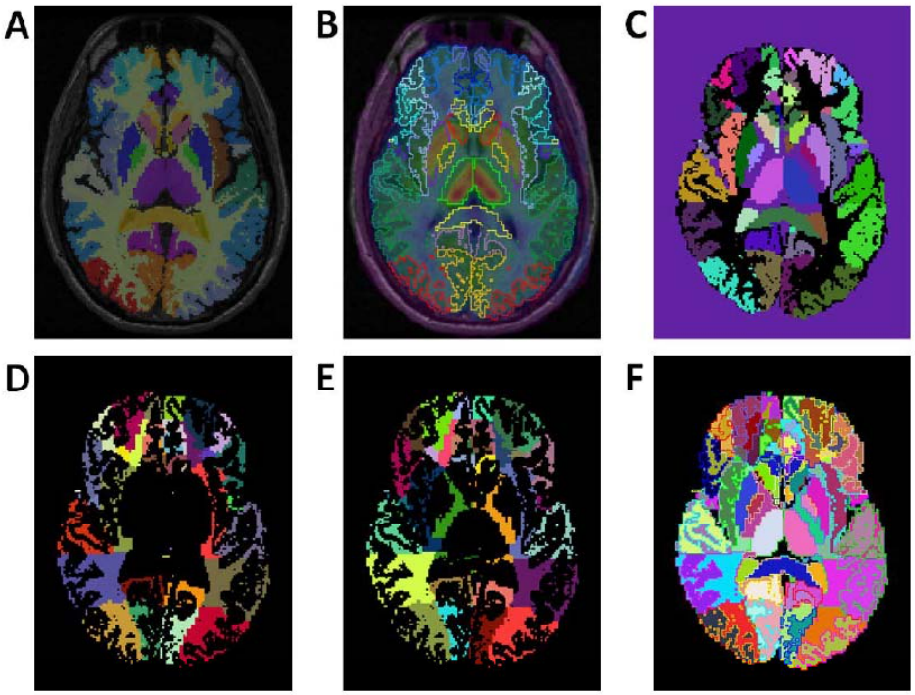
Table S4

Configuration	all subjects			Aß-positive subjects		
	HC (N = 88)	MCI (N = 148)	AD (N = 22)	HC (N = 26)	MCI (N = 66)	AD (N = 16)
2-year-Δ-COMP-SUVR-FULL$_{CSO}$	0.9±4.7%	1.1±5.2%	0.1±4.0%	0.1±4.9%	0.9±5.9%	-1.2±4.1%
2-year-Δ-COMP-SUVR-GM$_{CSO}$	1.7±4.9%	2.3±6.1%	1.4±4.9%	0.8±5.6%	1.8±6.7%	-0.8±5.8%
2-year-Δ-COMP-SUVR-PVEC$_{CSO}$	4.4±10.1%	5.5±10.7%	4.4±9.0%	3.2±11.2%	4.1±11.2%	0.4±12.7%
2-year-Δ-COMP-SUVR-FULL$_{WM\text{-}seg}$	0.0±3.0%	-0.1±2.5%	-0.8±2.2%	0.8±2.1%	0.2±2.0%	-0.8±2.2%
2-year-Δ-COMP-SUVR-GM$_{WM\text{-}seg}$	0.9±3.4%	1.2±3.2%	0.5±2.8%	1.9±2.3%	1.7±2.6%	0.9±2.4%
2-year-Δ-COMP-SUVR-PVEC$_{WM\text{-}seg}$	3.4±8.3%	4.1±8.0%	3.4±7.0%	4.9±4.9%	5.7±7.3%	4.6±5.0%

Mean (± SD) % changes over two years in [^{18}F]-AV45 uptake when assessed by the different COMP-SUVR methodological configurations (column 1) for the three diagnosis groups (columns 2 - 4) for all subjects. Columns 5 - 7 display corresponding results when only Aß positive subjects were considered.

Supplement 3

Figure S2: Image processing / PVEC illustrated on axial slices



(A) T1w-MRI fused with grey matter segments. Segments are obtained by masking the atlas with either a desired grey matter probability threshold (0.3 in our case) or a binary mask deriving from segmentation. Previously acquired transformations from PET to MRI and MRI to atlas/MNI space are used inversely to enable analyses in native PET or MRI space.

(B) MR fused with PET and overlaying with outlined VOIs.

(C) In the first step of preparation for VOI-based PVEC processing, white matter VOIs are removed, as those regions are subdivided into smaller regions in the subsequent steps. Second, the remaining original grey matter VOIs (multicolored) are subtracted from the atlas mask. This operation therefore defines voxels that were removed from the atlas during grey matter thresholding, or were assigned to white matter regions. The resulting voxels are illustrated in black.

(D) The mask created from voxels defined in the previous step is intersected with the atlas. This step creates complementary regions from the atlas definition.

(E) Finally, complementary white matter regions are created by adding voxels which were previously in white matter regions, but do not have a specific definition at this stage of processing. These regions are added by outlining to the original VOIs as complementary VOIs.

(F) Final regions are passed to the PVEC algorithm.

Send Orders for Reprints to reprints@benthamscience.ae

Current Alzheimer Research, 2016, *13*, 597-607

Impact of MRI-based Segmentation Artifacts on Amyloid- and FDG-PET Quantitation

Marcus Högenauer[#], Matthias Brendel[#], Andreas Delker, Sonja Därr, Mayo Weiss, Peter Bartenstein, Axel Rominger[*], for the Alzheimer's Disease Neuroimaging Initiative[**]

Department of Nuclear Medicine, University of Munich, Germany

Axel Rominger

Abstract: ***Introduction:*** Magnet resonance image (MRI)-based segmentations are widely used for clinical brain research, especially in conjunction with positron-emission-tomography (PET). Although artifacts due to segmentation errors arise commonly, the impact of these artifacts on PET quantitation has not yet been investigated systematically. Therefore, the aim of this study was to assess the effect of segmentation errors on [^{18}F]-AV45 and [^{18}F]-FDG PET quantitation, with and without correction for partial volume effects (PVE). ***Material and Methods:*** 119 subjects with both [^{18}F]-AV45, and [^{18}F]-FDG PET as well as T1-weighted MRI at baseline and at two-year follow-up were selected from the ADNI cohort, and their MRI brain images were segmented using PMOD 3.5. MRIs with segmentation artifacts were masked with the corresponding [^{18}F]-FDG PET standard-uptake-value (SUV) images to elucidate and quantify the impact of artifacts on PET analyses for six defined volumes-of-interest (VOI). Artifact volumes were calculated for each VOI, together with error-[%] and root-mean-square-errors (RMSE) in uncorrected and PVE corrected SUV results for the two PET tracers. We also assessed the bias in longitudinal PET data. ***Results:*** Artifacts occurred most frequently in the parietal cortex VOI. For [^{18}F]-AV45 and [^{18}F]-FDG PET, the percentage-errors were dependent on artifact volumes. PVEC SUVs were consequently more distorted than were their uncorrected counterparts. In static and longitudinal assessment, a small subgroup of subjects with large artifacts (≥1500 voxels; ≙5.06 cm³) accounted for much of the PET quantitation bias. ***Conclusion:*** Large segmentation artifacts need to be detected and resolved as they considerably bias PET quantitation, especially when PVEC is applied to PET data.

Keywords: Amyloid-PET, artifacts, FDG-PET, masking, MRI, PVEC, segmentation.

1. INTRODUCTION

In recent years, magnetic resonance imaging (MRI)-based segmentations of brain into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) are increasingly important for imaging studies in diverse disorders, including Alzheimer's disease (AD) [1, 2], multiple sclerosis [3], epilepsy [4] or schizophrenia [5, 6]. Moreover, molecular brain imaging techniques such as positron-emission-tomography (PET) [7] and single-photon-emission-computed-tomography (SPECT) [8] benefit from MRI-based segmentations, such that molecular changes can be investigated in specific target regions.

Indeed, MRI-based segmentations are essential for PET quantitation, especially in dementia studies entailing partial volume effect (PVE) corrected quantitation of amyloid [9] and/or [^{18}F]-fluorodesoxyglucose ([^{18}F]-FDG) [10, 11]. Although partial volume effect correction (PVEC) is also feasible, independent of MRI (*e.g.* by deconvolution), optimal results are obtained when using anatomically-oriented segmentation based on individual structural MRI [12]. As such, widely used PVEC methods such as the geometric transfer matrix (GTM) [13] or Müller-Gärtner methods [14] depend critically on correct segmentation of individual MR images. Volume artifacts due to segmentation errors arise commonly, with consequences for PET quantitation. This problem has led to intensive efforts aiming to enhance the existing automated segmentation methods such as Multi-Atlas Propagation and Segmentation [15], Bridge Burner [16], Brain Extraction Tool [17] and Hybrid Watershed Algorithm [18]. Despite considerable progress, entirely faultless segmentation procedures are not available; the impact of such segmentation artifacts on PET quantitation, and on error propagation with PVEC, has not yet been investigated systematically. Therefore, the aim of this study was to assess the effect of MRI segmentation errors on uncorrected and PVE-corrected [^{18}F]-AV45 and [^{18}F]-FDG PET quantitation in a large dataset of scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

*Address correspondence to this author at the Department of Nuclear Medicine, University of Munich, Germany; Tel: +49 (0)89 4400 74650; Fax: +49 (0)89 4400 77646; E-mail: axel.rominger@med.uni-muenchen.de

***Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/ uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*

[#]both authors contributed equally

1875-5828/16 $58.00+.00

2. MATERIAL AND METHODS

2.1. Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

2.2. Image Data Acquisition

2.2.1. ADNI [^{18}F]-AV45 PET and [^{18}F]-FDG PET Acquisition and Pre-Processing

The [^{18}F]-AV45 PET images had been acquired using Siemens, GE and Philips PET scanners according to a standard dynamic protocol with emission recording at 50-70 min following the intravenous injection of 370 ± 37 MBq [^{18}F]-AV45. [^{18}F]-FDG PET images had been acquired using Siemens, GE and Philips PET scanners according to one of three standard emission recording protocols (30–60 minute dynamic, 30–60 minute static, 0–60 minute dynamic) following the intravenous injection of 185 ± 19 MBq [^{18}F]-FDG. Data were corrected for both scatter and measured attenuation, which was determined using the CT scan for PET/CT scanners, or a transmission scan with [^{68}Ge] or [^{137}Cs] rotating rod sources for PET-only scanners. Images were reconstructed using scanner-specific algorithms, and sent to the University of Michigan, where they were reviewed for artifacts, anonymized, and transmitted to the Laboratory of Neuro Imaging (LONI) for storage. Further details are available in the ADNI PET technical procedures manual (http://www.adni-info.org/Scientists/Pdfs/ ADNI2_ PET_Tech_Manual_0142011.pdf).

Downloaded [^{18}F]-FDG PET images in DICOM format had been pre-processed in four steps: 1) motion correction by coregistration of single five minute frames; 2) time frame averaging (30-60 min p.i.); 3) coregistration of longitudinal data to the baseline scan and reorientation in a standardized 160x160x96 matrix consisting of 1.5 mm cubic voxels; 4) smoothing with a scanner-specific filter function to an isotropic resolution of 8 mm. Details can be found at (http://adni.loni.usc.edu/methods/pet-analysis/pre-processing/) [19, 20].

2.2.2. ADNI MRI Acquisition and Pre-Processing

T1-weighted (T1w) MRI scans had been acquired using Siemens, GE or Philips MRI scanners according to a standard protocol [21] involving acquisitions of two 3-D MPRAGE imaging sequences per subject. Of the two images acquired per subject and time-point, the ADNI quality assurance team selected the qualitatively better image for preprocessing.

MRI preprocessing involved: 1) application of a scanner-specific correction for gradient nonlinearity distortion (Gradwarp) [22]; 2) correction for image intensity non-uniformity (B1) [21]; 3) application of a histogram peak sharpening algorithm for bias field correction (N3) [23]; 4) application of spatial scaling factors obtained by phantom measurements. For images acquired on Philips scanners, B1 correction was already implemented, and the gradient systems with this instrument tended to be inherently linear [21].

2.3. Patient Selection and Study Design

138 subjects with both [^{18}F]-AV45 and [^{18}F]-FDG PET as well as T1w MPRAGE MRI at baseline (BL) and two-year follow-up (FU) were selected from ADNI-GO/2. Their clinical diagnoses at BL were healthy control (HC) (N=42), MCI (N=86) and AD (N=10). Pre-processed [^{18}F]-AV45 PET, [^{18}F]-FDG PET brain images and temporally corresponding T1w MPRAGE images were downloaded from the database as available on September 2013. After processing, all images were visually checked and MRI segmentation fails were identified as cases of complete mislabelling of the compartments (GM/WM/CSF) or misregistration resistant to manual adjustment. By these criteria, images of 13 subjects failed either at BL or FU, while images of six subjects failed at both BL and FU. Thus, a total of 25/276 (9.1%) failed images were excluded from further analysis, corresponding to a total of 19 subjects (HC: N=6, MCI: N=12, AD: N=1) (Fig. 1).

2.4. Image Analyses

2.4.1. Coregistration and Segmentation

[^{18}F]-AV45 and [^{18}F]-FDG PET images were rigidly coregistered to their individual MRI to obtain a linear PET-to-MRI transformation in PMOD. Non-rigid coregistration of individual MRIs to the standard Montreal Neurological Institute (MNI)-space was used to obtain a nonlinear MRI-to-MNI transformation. Next, the concatenation of PET-to-MRI and MRI-to-MNI transformations was used to transform the PET images in MNI-space. Meanwhile, the T1w MRIs were segmented to GM, WM and CSF compartments [24]. To this end we implemented the segmentation algorithm with the following parameters: Sampling distance,

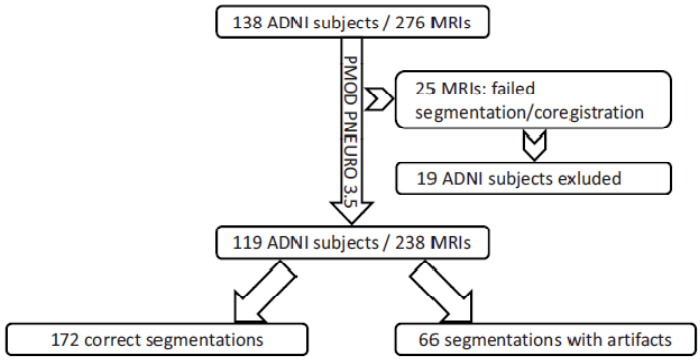


Fig. (1). Schematic overview of study design and subject selection. Image data from 138 ADNI subjects, undergoing baseline and two-year follow-up MRIs, [^{18}F]-AV45 PET and [^{18}F]-FDG PET, were processed by PMOD PNEURO 3.5. Subsequent visual identification of completely failed segmentation or coregistration resulted in exclusion of 19 subjects, and a final cohort of 238 MRIs totalling 119 subjects (HC: N=36, MCI: N=74, AD: N=9) for further analysis. 172 artifact-free segmentations and 66 segmentations with at least one artifact were obtained.

which determined the density of voxels considered in the calculation, was used at a standard 6 mm setting. Moderate bias regularisation served to compensate for variations of the image intensity across the field-of-view, and a moderate clean-up procedure rectified the incorrect segmentation along the compartment boundaries. To guarantee the best possible uniformity of segmentations, all parameters were used unadjusted in standardized manner, even when some mislabelling consequently occurred. In such cases, the segmentations were excluded, as described above. Volumes-of-interest (VOIs) of 83 cortex areas from the Hammers atlas were defined in individual GM templates for each subject in the MNI-space [25]. All image MR processings were performed with the PMOD PNEURO 3.5 tool.

2.4.2 Artifact Assessment and Masking

BL and FU segmentations of 119 subjects (HC: N=36; MCI: N=74; AD: N=9) were visually checked slice-by-slice for artifacts. In cases when an artifact was visually identified, as occurred for about 20 voxels per plane, the segmentation was masked by the individual [^{18}F]-FDG PET in order to exclude the artifact voxels. The mask binarization of [^{18}F]-FDG PET was performed by an individual intensity threshold (20-37% of total activity) optimally fitting the MRI brain border, in order to guarantee exclusion of all artifact voxels. In cases where these requirements could not be fulfilled, as occurred for <5% of all [^{18}F]-FDG PET images, the threshold was selected, which came most close to rescuing the assignment of all real brain voxels. Subsequently, the remaining artifact regions were cropped by setting to zero voxels outside the binary mask (Fig. **2**). For further analyses, the corrected MRI segmentations without artifacts were used for processing of the corresponding [^{18}F]-AV45 and [^{18}F]-FDG PET in PMOD PNEURO 3.5 tool again.

2.4.3. VOI-based PVEC

For each VOI defined in the Hammers atlas [25], we calculated PVE-corrected values [13, 26, 27]. The algorithm included an extra-cerebral background VOI implemented in the PNEURO 3.5 tool. We used a GM threshold 0.3 and a full-width-at-half-maximum (FWHM) defined of 8 mm; details of the PVEC algorithm are provided in Supplement **1**.

2.4.4. Analyses of Artifact Volumes and PET-VOI Data

Six GM VOIs were defined by combining bilateral Hammers atlas VOIs, i.e. frontal, parietal, temporal, and posterior cingulate cortex (PCC), a composite of all these cortical regions, as well as an entire cerebellum GM VOI. Uncorrected and PVE-corrected standard-uptake-values (SUV) were calculated for [^{18}F]-AV45 and [^{18}F]-FDG PET using non-masked (SUV_{art}) and masked (SUV_{true}) VOI sets. With these data the following analyses were performed:

2.4.4.1. Artifact Volume Quantitation

Artifact sizes were quantified by subtracting cropped from uncropped volumes. A volume threshold of </≥5.06 cm³ (≙1500 voxels) was defined for further subanalyses. Artifact volumes were estimated by a hot 3D VOI tool and 1500 voxels were the approximated lower bound of these artifacts volumes.

2.4.4.2. Error-[%] in PET SUVs and Correlation with Artifact Volumes

For each of the six VOIs, we calculated error-[%] in the uncorrected and PVE-corrected SUVs according to equation 1. Error-[%] in SUVs were then correlated with artifact volumes.

$$Error\text{-}[\%] = \frac{SUV_{art} - SUV_{true}}{SUV_{true}} \times 100 \qquad \text{Eq. 1}$$

2.4.4.3. Root-mean-square-error (RMSE)

RMSE were calculated for uncorrected and PVE-corrected SUVs according to equation 2:

$$RMSE\text{-}[\%] = \frac{\sqrt{\sum_{i=1}^{n}(SUV_{art} - SUV_{true})^2}}{\sqrt{\sum_{i=1}^{n}(SUV_{true})^2}} \times 100 \qquad \text{Eq. 2:}$$

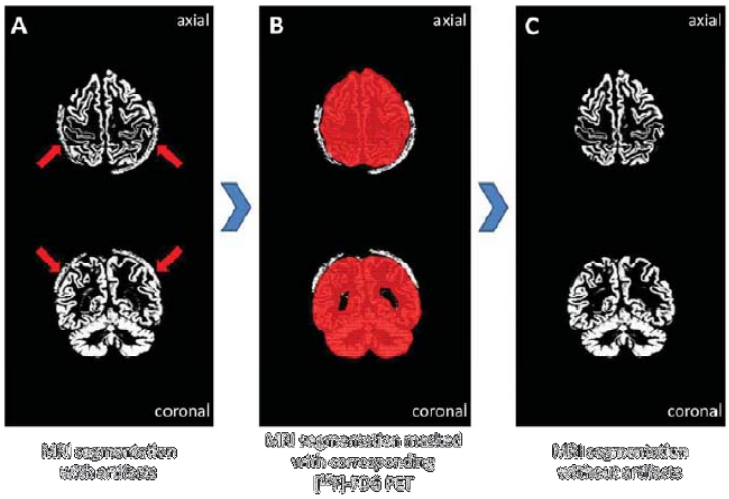


Fig. (2). Workflow of artifact masking. **(A)** Example of a segmentation artifact (highlighted by arrows) in axial and coronal plane. **(B)** Binarization of the corresponding [^{18}F]-FDG PET was performed by the individual threshold (20-37% of total activity) give the best fit to the MRI brain border. **(C)** Resulting artifact-free MRI segmentation after cropping process.

2.4.4.4. Longitudinal PET Assessment

Distribution of artifacts relatively to the different brain regions was assessed in the BL and FU study. Percentage changes in longitudinal SUVs (Δ-SUV-[%]) between BL and FU were calculated for the SUV data with artifacts (equation 3) or MR-artifact-free (equation 4) PET results.

$$\Delta\text{-}SUV\text{-}[\%]_{art} = \frac{SUV_{art}FU - SUV_{art}BL}{SUV_{art}BL} \times 100 \quad \text{Eq. 3}$$

$$\Delta\text{-}SUV\text{-}[\%]_{true} = \frac{SUV_{true}FU - SUV_{true}BL}{SUV_{true}BL} \times 100 \quad \text{Eq. 4}$$

The consecutive resultant bias in longitudinal PET data was calculated by Equation 5:

$$Bias\text{-}\Delta\text{-}SUV\text{-}[\%] = \Delta\text{-}SUV\text{-}[\%]_{true} - \Delta\text{-}SUV\text{-}[\%]_{art} \quad \text{Eq. 5}$$

2.4.5. Semi-automatic Masking Approach

In order to investigate a possibility method for facilitating and simplifying the masking process, we examined template masks deriving from the complete data set. In this approach, the masked segmentations from all 238 individual sets of VOIs deemed without artifacts were binarized as unity for all voxels included in any VOI. The 238 binarized images were then averaged to obtain a probability value for each voxel of "inclusion in the VOI-set" within the whole population, as has been done previously for probabilistic atlas generation [28]. Standardized masks were generated with either 100% (100%-voxel-mask) or 50% (50%-voxel-mask) probability level threshold. Thus, the 100%-voxel-mask contained every voxel of all VOI sets, while the much more stringent 50%-voxel-mask contained all voxels present in the VOI-set of at least half of the subjects. These masks were subsequently used to test the statistical template approach in the subjects for whom there were more than 1500 artifact voxels.

2.4.6. Statistics

Means (± SD) and ranges of volume as well as error-[%] in uncorrected and PVE-corrected SUVs were assessed for both tracers for all six VOI regions. Additionally Pearson's coefficient of correlation (r) was assessed for uncorrected and PVE-corrected SUV error-[%] with volumes of artifacts. Differences between uncorrected and PVE-corrected arising from MR segmentation artifacts were assessed by comparing single square errors with a paired student's *t*-test. Furthermore distinctions between different mask settings (raw, 50%- and 100%-voxel-mask) were as well assessed by a paired students *t*-test. The null hypothesis was rejected for $p<0.05$. In the longitudinal PET assessments the moduli of the mean (± SD) for Bias-Δ-SUV-[%] and Bias-Δ-VOL-[cm^3] were calculated and reported with the respective ranges.

3. RESULTS

3.1. Demographics

Detailed demographics of the study cohort are provided in Table **1**.

3.2. Artifact Volume Assessment and Error-[%] in SUV

In 66/238 (27.7%) of the MRIs, at least one artifact was identified by visual inspection, leading to a total of 276 artifacts, which were most frequent in the parietal region (N=84). Details of artifact volumes of all six assessed regions are presented in Table **2**, and the highest artifact volumes were found in parietal (mean: 2.9 ± 5.6 cm^3; range: **0** - 26.7 cm^3) and temporal regions (mean: 2.4 ± 2.9 cm^3; range:

Table 1. Demographics of healthy controls (HC), patients with mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD).

Diagnosis	N	Age	Gender
(Baseline)		(y ± SD)	(f / m)
HC	36	77.4 ± 6.4	13 (36%) f / 23 (64%) m
MCI	74	70.2 ± 7.9	33 (45%) f / 41 (55%) m
AD	9	73.4 ± 8.2	6 (67%) f / 3 (33%) m

Table 2. Mean (±SD) and ranges (MIN / MAX) of artifact volumes, as well as counts (N) of artifacts in different brain regions. Correlation coefficients (r) are given for the relationship between artifact volumes and the error-[%] in SUV of [^{18}F]-AV45 and [^{18}F]-FDG PET, for uncorrected and PVE-corrected data.

Region	Artifact			Error-[%] in [^{18}F]-AV45 PET						Error-[%] in [^{18}F]-FDG PET					
	Volume [cm^3]			uncorrected			PVEC			uncorrected			PVEC		
	MEAN ± SD	MIN / MAX	N	MEAN ± SD	MIN / MAX	r	MEAN ± SD	MIN / MAX	r	MEAN ± SD	MIN / MAX	r	MEAN ± SD	MIN / MAX	r
Frontal	0.9 ± 1.5	0.0 / 11.5	51	-0.1 ± 0.5	-4.1 / 0.4	-0.88	-0.5 ± 1.2	-9.2 / 0.2	-0.94	-0.3 ± 0.6	-4.7 / 0.0	-0.99	-0.2 ± 0.9	-6.8 / 0.5	-0.88
Parietal	2.9 ± 5.6	0.0 / 26.7	84	-1.2 ± 2.8	-14.8 / 0.4	-0.94	-3.0 ± 7.0	-36.6 / 0.8	-0.90	-1.9 ± 3.7	-17.9 / 0.0	-0.99	-2.4 ± 5.9	-28.7 / 0.4	-0.98
Temporal	2.4 ± 2.9	0.0 / 17.5	67	-0.4 ± 0.8	-6.3 / 0.4	-0.82	-0.6 ± 1.3	-9.6 / 1.3	-0.68	-0.8 ± 1.2	-8.5 / 0.0	-0.96	-0.6 ± 1.5	-11.1 / 0.3	-0.84
PCC	0.1 ± 0.2	0.0 / 0.2	5	0.0 ± 0.0	-0.1 / 0.1	-0.25	0.4 ± 1.3	-0.5 / 6.6	-0.07	0.0 ± 0.0	-0.1 / 0.0	-0.94	0.4 ± 1.0	-0.1 / 5.6	-0.03
Composite	6.2 ± 8.2	0.1 / 38.9	207	-0.5 ± 1.0	-5.2 / 0.3	-0.89	-1.1 ± 2.2	-10.3 / 0.1	-0.84	-0.9 ± 1.3	-6.3 / 0.0	-0.98	-0.9 ± 1.9	-9.2 / 0.2	-0.94
Cerebellum	1.4 ± 1.3	0.0 / 5.4	69	-0.3 ± 0.5	-2.1 / 0.1	-0.82	-0.5 ± 0.9	-3.8 / 1.4	-0.76	-0.7 ± 0.6	-2.8 / 0.0	-0.98	-0.5 ± 0.6	-2.6 / 0.1	-0.90

0 - 17.5 cm^3). Volumes for PCC were negligibly affected by segmentation artifacts.

Strongest correlations of artifact volumes with error-[%] in SUV were detected in the parietal cortex region for [^{18}F]-AV45 PET (uncorrected: r = -0.94; PVEC: r = -0.90) (Fig. **3A**) as well as for [^{18}F]-FDG PET (uncorrected: r = -0.99; PVEC: r = -0.98) (Fig. **3B**). Considerable error-[%] in SUV were detected above the defined artifact volume threshold of 5.06 cm^3 (≙1500 voxels), which was present in 9/66 segmentations with artifacts. However there was no relationship between artifact sizes and clinical diagnoses.

3.3. RMSE-[%] in uncorrected and PVE-corrected PET SUVs

Highest RMSE-[%] in SUV was observed in the parietal region. For both tracers, PVE-corrected data was significantly more affected by artifacts when compared to uncorrected SUVs ($p<0.05$), as illustrated in (Fig. **4A/B**), *e.g.* in the parietal region ([^{18}F]-AV45 PET: 7.1% *vs.* 2.6% / [^{18}F]-FDG PET: 6.3% *vs.* 4.2%). Furthermore, RMSE-[%] results were analysed for subgroups with </≥1500 artifact voxels (57 segmentations with <1500 artifact voxels and 9 segmentations with ≥1500 artifact voxels). The small group with artifacts of ≥1500 voxels nonetheless accounted for the largest share of total RMSE-[%] in SUV for both tracers (Fig. **4C**), while RMSE-[%] in all regions were consistently below 1.5% for the subgroup with <1500 voxels, irrespective of radiotracer or whether or not PVEC was applied.

3.4. Distribution of Artifacts and Impact on Longitudinal PET Studies

The 66 segmentations with artifacts were obtained from 39 different subjects (27 of whom had artifacts in their MR-segmentations both at BL and FU). The latter group comprised a regional concordance of 88/112 (78.6%) artifacts at both MR assessments. In the longitudinal PET measure of Bias-Δ-SUV-[%], again a small fraction of subjects (Bias-Δ-VOL-[cm^3] ≥1500 voxels) had the highest impact on PET quantitation, which increased as a function of artifact size (Fig. **5 A,B**).

Highest Bias-Δ-SUV-[%] of up to 36% was found in parietal region of PVE-corrected [^{18}F]-AV45 PET (Table **3**). For both PET tracers, longitudinal assessments were more affected by MR segmentation artifacts in PVE-corrected SUVs than for uncorrected SUVs ([^{18}F]-AV45: $p<0.001$ / [^{18}F]-FDG: $p<0.05$). The group of 34 subjects with a Bias-Δ-VOL-[cm^3] <1500 voxels revealed a negligible mean Bias-Δ-SUV-[%] to FU even in the most affected region, i.e. parietal

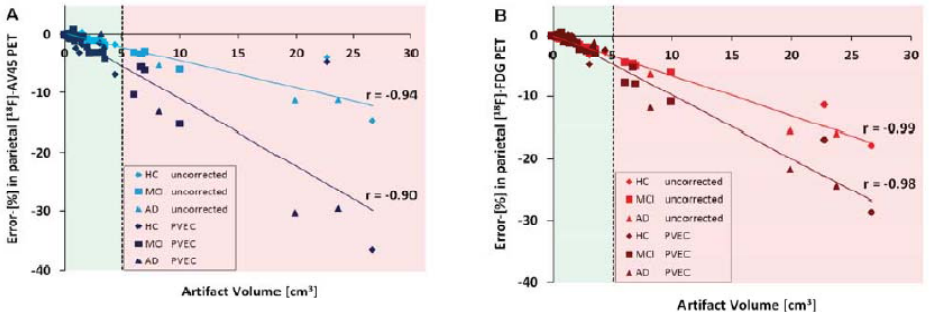


Fig. (3). **(A)** Correlation of artifact volumes [cm^3] with error-[%] in uncorrected (light blue) and PVEC (dark blue) for [^{18}F]-AV45 PET in parietal cortex. **(B)** Correlation of artifact volumes [cm^3] with error [%] in uncorrected (light red) and PVEC (dark red) for [^{18}F] FDG PET in parietal cortex. In both charts, single subjects are depicted by diamond (HC), square (MCI), and triangle (AD). Pearson's correlation coefficients (r) are provided. Dotted line represents the threshold above 5.06 cm^3 (≙1500 voxels, red area), which had considerable impact on PET quantitation for both tracers.

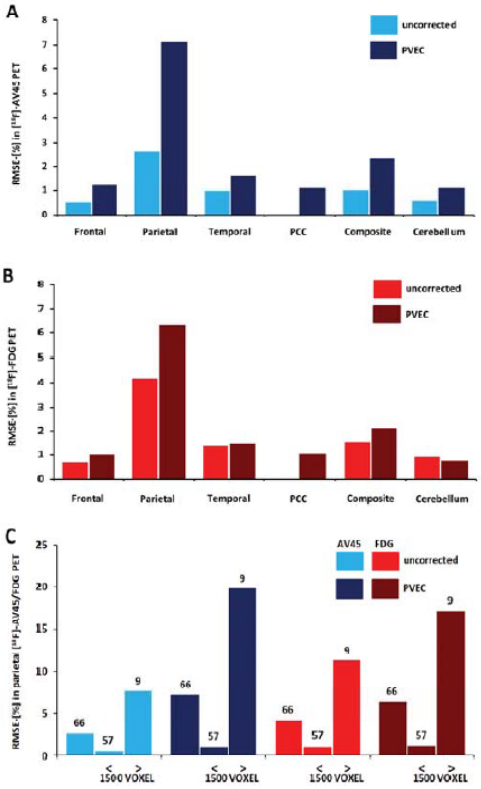


Fig. (4). **(A)** RMSE-[%] in [^{18}F]-AV45 PET SUV caused by artifacts for uncorrected (light blue) and PVE-corrected (dark blue) values. **(B)** RMSE-[%] in [^{18}F]-FDG PET SUV caused by artifacts for uncorrected (light red) and PVE-corrected (dark red) values. **(C)** RMSE-[%] of uncorrected (light blue/red) and PVEC (dark blue/red) values of the parietal cortex VOI before (N=66) and after stratifying in subgroups containing <1500 (N=57) and ≥1500 (N=9) artifact voxels.

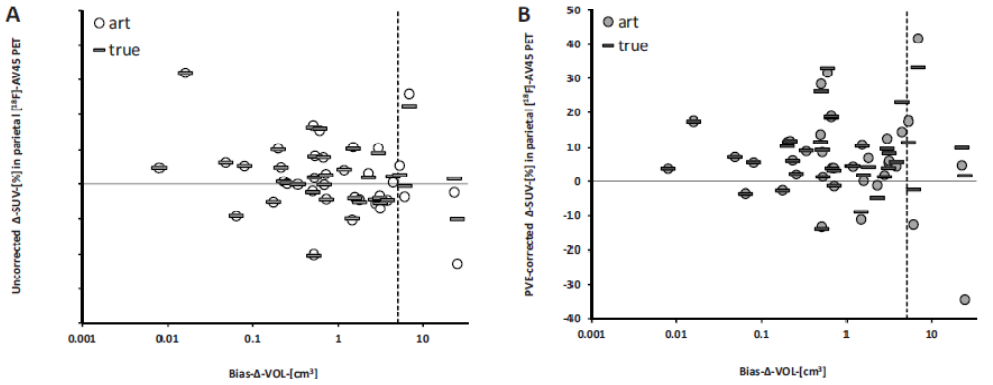


Fig. (5). **(A)** Discrepancies in two year follow-up PET results of single subjects as expressed by true (light grey bars) and biased (white circles) Δ -SUV-[%] (uncorrected [^{18}F]-AV45 PET) are presented as a logarithmic function of longitudinal apparent MR-volume differences (Bias-Δ-VOL-[cm^3]). **(B)** Discrepancies in two year follow-up of PET results of single subjects as expressed by true (dark grey bars) and biased (light grey circles) Δ-SUV-[%] (PVE-corrected [^{18}F]-AV45 PET) are presented as a logarithmic function of longitudinal apparent MR volume differences (Bias-Δ-VOL-[cm^3]). Dotted line represents the cut-off threshold of 5.06 cm^3 (≙1500 voxels). Parietal cortex region serves as the illustrative example.

Table 3. **Mean (±SD; modulus) and MIN / MAX for two-year longitudinal Bias-Δ-VOL-[cm^3] as well as Bias-Δ-SUV-[%] of [^{18}F]-AV45 and [^{18}F]-FDG. Results are given separately for uncorrected and PVE-corrected Bias-Δ-SUV-[%]. Significant differences between uncorrected and PVE-corrected data are indicated by * p < 0.05; ** p < 0.01; *** p < 0.001.**

Region	Bias-Δ-VOL-[cm^3]		Bias-Δ-SUV-[%] in [^{18}F]-AV45 PET				Bias-Δ-SUV-[%] in [^{18}F]-FDG PET			
			uncorrected		PVEC		uncorrected		PVEC	
	MEAN ± SD	MIN / MAX	MEAN ± SD	MIN / MAX	MEAN ± SD	MIN / MAX	MEAN ± SD	MIN / MAX	MEAN ± SD	MIN / MAX
Frontal	0.8 ± 1.4	-2.6 / 8.5	0.2 ± 0.5	-3.0 / 0.7	0.4 ± 0.9**	-5.9 / 1.4	0.3 ± 0.6	-4.1 / 0.8	0.3 ± 0.9	-5.8 / 0.6
Parietal	2.6 ± 5.2	-24.7 / 6.9	1.0 ± 2.1	-3.5 / 12.6	2.7 ± 6.0*	-8.5 / 35.9	1.6 ± 3.2	-4.9 / 17.3	2.4 ± 5.4	-7.7 / 30.2
Temporal	2.3 ± 3.0	-11.5 / 13.6	0.4 ± 0.8	-5.0 / 1.8	0.7 ± 1.2**	-7.3 / 2.2	0.8 ± 1.3	-7.2 / 3.2	0.7 ± 1.6	-9.9 / 2.9
PCC	0.1 ± 0.2	0.0 / 0.2	0.0 ± 0.0	-0.1 / 0.1	0.4 ± 0.9**	-5.5 / 1.4	0.0 ± 0.0	0.0 / 0.2	0.4 ± 0.8**	-4.3 / 1.9
Composite	5.7 ± 7.6	-36.5 / 23.9	0.4 ± 0.7	-3.2 / 3.5	0.9 ± 1.7**	-5.4 / 9.3	0.8 ± 1.1	-4.0 / 4.7	0.9 ± 1.6	-5.8 / 7.2
Cerebellum	1.3 ± 1.2	-4.9 / 4.1	0.3 ± 0.4	-1.0 / 1.9	0.7 ± 0.9***	-3.1 / 4.1	0.6 ± 0.5	-2.1 / 2.2	0.5 ± 0.5**	-2.1 / 2.1

cortex: [^{18}F]-AV45 uncorrected: -0.02%; PVE-corrected: -0.02%; [^{18}F]-FDG uncorrected: +0.09%; PVE-corrected: -0.16%.

3.5. Semi-automatic Masking Approach

By application of the statistically defined 100%-voxel-mask, only minor effects on artifact volumes were observed, whereas the more stringent 50%-voxel-mask led to a considerable reduction of mean parietal cortex segmentation artifact volumes (14.1 cm^3 vs. 3.4; p<0.001). However, the 50%-voxel-mask resulted in distinct underestimation of true brain voxels in the frontal region (-7.6 cm^3) (Fig. **6A**). Despite the reduction of parietal cortex artifact volumes, SUV RMSE-[%], were reduced by only 50% in uncorrected and PVE-corrected PET of both tracers (Fig. **6B**) when using the strict 50%-voxel-mask.

4. DISCUSSION

This is the first systematic investigation of the impact of MRI segmentation-related artifacts on the quantitation of brain PET data. Longitudinal [^{18}F]-AV45 and [^{18}F]-FDG PET recordings from the ADNI database were analyzed according to uptake in GM VOIs derived from MRI segmentation, with and without correction for segmentation artifacts.

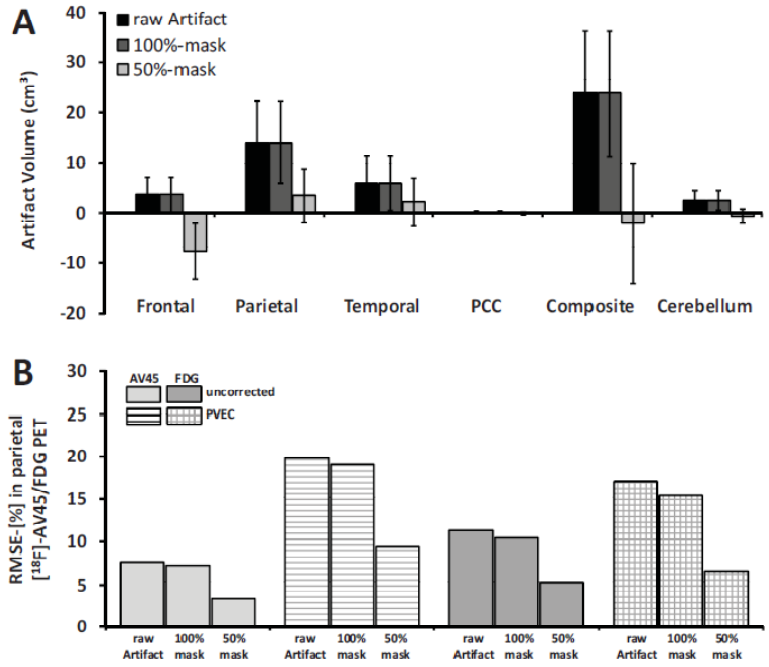


Fig. (6). **(A)** Mean (±SD) artifact volumes either for raw (black bars) segmentations or after standardized masking by the 100% probabilistic-mask (darker grey bars) and the more stringent 50%-mask (lighter grey bars) for six different brain regions (expressed for the nine MR segmentations with extremely large artifacts (≥1500 voxels)).**(B)** RMSE-[%] in [^{18}F]-AV45 and [^{18}F]-FDG PET SUV for the nine MR segmentations with extremely large artifacts (≥1500 voxels). RMSE-[%] for the parietal cortex region are given either for raw segmentations or after standardized masking by the 100%-mask and 50%-mask for uncorrected and PVE-corrected values.

MRI-based segmentation artifacts propagated to bias in PET quantitation in proportion to the volume of segmentation errors; this effect was most pronounced in PVE-corrected data. There was considerable variance in the magnitude of bias in individual SUVs, with substantial impact on the group results arising from the few subjects with especially large artifacts; identification of such patients is required in order to avoid biased PET quantitation in multicenter cohort studies.

4.1. Artifact Assessment and Masking

We found that one quarter of processed segmentations revealed at least one artifact by visual quality control, indicating a high frequency of relevant segmentation artifacts in this multi-center collective. The parietal cortex region proved to be most vulnerable to artifacts, most likely due to inherently greater likelihood of bridge building VOIs adjacent to this brain area. This is concordant with previous studies showing the meninges and large sinuses, which are notably present in vicinity to the parietal region, were the non-brain structures most often falsely segmented in brain MRI [16]. The rather sophisticated masking method used in this investigation afforded identification of the optimal threshold for masking of VOIs. Through careful editing, nearly perfect masking of artifacts was achievable, in a trade-off between the most rigorous possible exclusion of all artifact voxels, and preservation of all real cortical voxels in the VOI. As an alternative, we also tested a semiautomatic masking approach, but this proved without benefit, as artifact voxels were not sufficiently excluded by a moderate stringency threshold mask, or bias was introduced by elimination of real cortical voxels when a strict threshold mask was applied; we found that RMSE-[%] in the PET result was reduced by only 50% when using the strict mask. This reflects the location of most artifacts in the direct vicinity to the brain, which brings a risk of overlap between true GM and artifact voxels in the MNI space when considering subjects with differing brain shape and atrophy. Therefore, we conclude that standardized semiautomatic masks are no substitute for individual masking to resolve segmentation artifacts by post-processing. Correct masking is aided by access to the individual [^{18}F]-FDG PET image, which is not available in many brain PET studies. When [^{18}F]-FDG PET is not available, masking is also possible through time-consuming manual identification and erasing of artifact voxels using VOI or ROI post-processing tools.

Despite the presence of extra-cerebral artifacts, overestimation of GM compared to WM (mis-segmentation) is well-

known to arise in MRI segmentation algorithms. This was examined in detail by Gutierrez and colleagues in [^{18}F]-FDG PETs of 19 patients with dementia, and they concluded that statistical parametric mapping (SPM) 5, which was also implemented in the PMOD algorithm used in the current study, provides reliable results, superior to those obtained with SPM 99 or SPM 2 [29]. Nonetheless, we found 9% completely failed MRI segmentations, which is also in line with our previous findings in a larger ADNI dataset of 1300 subjects [30]; this could reflect the challenge presented by atrophy in the aged or disease brain relative to the MNI target, which represents healthy young subjects.

4.2. Influence of Segmentation Artifacts on PET Quantitation

Bias in PET quantitation correlated highly with the individual volume of artifacts. This is compelling, as artifact voxels are predominantly located outside the cerebral cortex, and thus have very low radiotracer uptake, which consequently exert a large bias on the apparent VOI activity. Therefore, spuriously low SUV values are the logical consequence of MR segmentation error, and are most pronounced when brain tracer uptake is relatively high compared to the extracerebral background activity.

One objective of this study was to establish the size of artifacts leading to relevant errors in PET quantitation, with likely consequences for between-group comparisons. This is a matter of interest as the extensive quality control performed in this investigation is labour intensive, and might not be feasible in planned studies with many hundreds of subjects. We found artifacts comprising <1500 voxels to be of minor relevance in PET quantitation at a single time point, whereas artifacts larger than 1500 voxels (5 cm^3) in specific brain VOIs had a considerable impact on PET quantitation. Thus, it is very important to take into account the target region sizes of PET studies, as quantitation in small brain regions is inherently more vulnerable to MR segmentation artifacts. Thus for example, amyloid-PET [^{18}F]-AV45 will scarcely be affected if the end point is semi-quantitative assessment of amyloid positivity or negativity within the entire cortical GM VOI, which comprises 143,000 voxels. In contrast, [^{18}F]-FDG PET analysis in a parietal subregion comprising as few as 4,500 voxels is highly vulnerable to the presence of a large artifact.

From our experience, all cases with segmentation artifacts exceeding the 1500 voxel threshold were easily discerned by simply scrolling through the axial layer (< 1 min), which should enable rapid screening of segmentations requiring correction.

4.3. Impact of Segmentation Artifacts on PVEC

Accounting for effects of atrophy is crucial in brain PET studies of healthy aging and AD, as the loss of cortical volume will bias observations of metabolism and ligand binding due to increasing PVEC (12). Particular implications of PVEC on [^{18}F]-flutemetamol/amyloid PET quantitation were recently assessed in a phantom study employing a new region-based voxel-wise technique, in comparison to well-established VOI-based methods [9]. The authors concluded that PVEC enhances precision of regional amyloid PET quantitation, and should be considered as obligatory in clinical trials with PET endpoints. In the present work, we found that PET quantitation with PVEC was more affected (than was native, uncorrected PET data) by segmentation artifacts, regardless of the subregion or the radiotracer. This is expected from the methodological principles of VOI-based PVEC, since the presence of (extracerebral) artifacts will tend to create the appearance of hypertrophy in the GM. In this circumstance, a spuriously low correction PVEC is performed by the algorithm in artifact-contaminated regions, which was reflected by the greater underestimation of PET SUV values with PVEC when compared to uncorrected data (Figs. **3 A, B**). This result is in line with findings in previous works, where (other than misregistration) mis-segmentation (over/underestimation of GM) had the most pronounced impact on region-based [31] and/or voxel-based [32, 33] PVEC accuracy. Efforts towards detection and elimination of segmentation artifacts are consequentially justified when PVEC is implemented in the data analysis.

4.4. Longitudinal Studies

Monitoring of therapeutic trials in neurodegenerative diseases with well-established non-invasive biomarkers like MRI and [^{18}F]-FDG PET is of burgeoning importance [34]. However, these longitudinal investigations tend to suffer from low effect sizes magnitudes, even for therapeutic interventions lasting for more than two years of disease progression. This is exemplified in a multimodal image study of a trial of memantine in mild to moderate AD [35]; in the course of one year, cerebral glucose metabolism decreased by a mean of only -1.8% in memantine-treated patients versus -3.0% in the placebo group. It is self-evident that a longitudinal single subject bias (over two years' follow-up) of 17% in the uncorrected [^{18}F]-FDG PET uptake, as we observed in parietal cortex, the region most vulnerable to segmentation artifact of our study, would detrimentally influence such results, if not resolved or excluded. Nevertheless, the overall impact of smaller artifacts should normally be compensated by sufficient sample sizes; we found the mean Bias-Δ-SUV-[%] in [^{18}F]-FDG PET to be less than 0.2% (for the highest affected parietal region) in the sub-group of 34 subjects with a Bias-Δ-VOL-[cm^3] <1500 voxels.

We found a high concordance of regional artifact distribution to baseline and follow-up segmentations in the two-year longitudinal setting (79%). Thus, we conclude that individual brain anatomy contributes to the occurrence of artifacts. We did not systematically investigate the possible influence of specific imaging sites on the magnitude of artifacts in this multi-center study, as we processed MRIs only when dual tracer PET imaging was also present. We did note that some sites tended to produce more MRIs that were affected by segmentation artifacts in consecutive post-processing (max. 50%; 10/20), but this effect of the site was not statistically significant.

4.5. Limitations

In the interests of clarity, we restricted our investigation only to VOI-based PVEC as implementing another PVEC approach (*e.g.* Müller-Gärtner method). However, pilot data (not shown) suggested that the effects of artifacts on voxel-

wise MRI-based PVEC tended to mirror our findings with VOI-based PVEC. This is unsurprising, as the bias of enlarged GM volumes remains constant for both approaches. The aim of this study was rather to emphasize the direct impact and consequences of MR segmentation artifacts in commonly used multi-center data by application of a standard (and widely used) PMOD algorithm. However, segmentation of other MRI sequences and use of different segmentation algorithms need to be explored in more detail. More sophisticated segmentation approaches would potentially lower the frequency of relevant artifacts, as specialized artifact-reducing features are increasingly implemented in those algorithms (*e.g.* bridge burner).

CONCLUSION

MRI-based segmentation artifacts have a size-dependent effect on PET quantitation in this multitracer follow-up study of HC, MCI and AD subjects. Individual SUVs were affected to a highly variable extent, with very substantial impact in those few subjects with especially large artifacts. When applying PVEC algorithms, the impact is even more pronounced than for uncorrected studies. Screening and editing of suspect MR segmentations is necessary to avoid biased PET quantitation. If available, [^{18}F]-FDG-PET is suitable for masking segmentation-based artifacts in such cases, whereas a standardized template based masking approach was inferior.

ABBREVIATIONS

AD	=	Alzheimer's disease
ADNI	=	Alzheimer's disease neuroimaging initiative
BL	=	baseline
CSF	=	cerebrospinal fluid
FDA	=	food and drug administration
[^{18}F]-FDG	=	[^{18}F]-fluorodesoxyglucose
FU	=	follow-up
FWHM	=	full-width-at-half-maximum
GTM	=	geometric transfer matrix
GM	=	grey matter
HC	=	healthy control
LONI	=	laboratory of neuro imaging
MCI	=	mild cognitive impairment
MNI	=	montreal neurological institute
MRI	=	magnetic resonance imaging
NIA	=	national institute on aging
NIBIB	=	national institute of biomedical imaging and bioengineering
PCC	=	posterior cingulate cortex
PET	=	positron-emission-tomography
PVE(C)	=	partial volume effect (correction)
RMSE	=	root-mean-square-error
SPECT	=	single-photon-emission-computed-tomography
SPM	=	statistical parametric mapping
SUV	=	standard-uptake-value
Δ-SUV-[%]	=	change in longitudinal SUV
T1w	=	T1-weighted
VOI	=	volume-of-interest
VOL	=	volume
WM	=	white matter

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

CONFLICT OF INTEREST

M.H., M.B., A.D., S.D., M.W., P.B. and A.R. report no disclosures.

ACKNOWLEDGEMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; ; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. We acknowledge profession textual revisions by Inglewood Biomedical Editing.

REFERENCES

[1] Anderson VM, Schott JM, Bartlett JW, Leung KK, Miller DH, Fox NC. Gray matter atrophy rate as a marker of disease progression in AD. Neurobiol Aging 33(7): 1194-202 (2012).

[2] Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, *et al*. 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. Brain 130(Pt 7): 1777-86 (2007).

[3] Sepulcre J, Sastre-Garriga J, Cercignani M, Ingle GT, Miller DH, Thompson AJ. Regional gray matter atrophy in early primary progressive multiple sclerosis: a voxel-based morphometry study. Arch Neurol 63(8): 1175-80 (2006).

[4] Ciumas C, Savic I. Structural changes in patients with primary generalized tonic and clonic seizures. Neurology 67(4): 683-6 (2006).

[5] Hutcheson NL, Clark DG, Bolding MS, White DM, Lahti AC. Basal ganglia volume in unmedicated patients with schizophrenia is associated with treatment response to antipsychotic medication. Psychiatr Res 221(1): 6-12 (2014).

[6] Oertel-Knochel V, Knochel C, Matura S, Rotarska-Jagiela A, Magerkurth J, Prvulovic D, *et al*. Cortical-basal ganglia imbalance in schizophrenia patients and unaffected first-degree relatives. Schizophr Res 138(2-3): 120-7 (2012).

[7] Kawachi T, Ishii K, Sakamoto S, Sasaki M, Mori T, Yamashita F, *et al*. Comparison of the diagnostic performance of FDG-PET and VBM-MRI in very mild Alzheimer's disease. Eur J Nucl Med Mol Imaging 33(7): 801-9 (2006).

[8] Goto H, Ishii K, Uemura T, Miyamoto N, Yoshikawa T, Shimada K, *et al*. Differential diagnosis of dementia with Lewy Bodies and Alzheimer Disease using combined MR imaging and brain perfusion single-photon emission tomography. AJNR Am J Neuroradiol 31(4): 720-5 (2010).

[9] Thomas BA, Erlandsson K, Modat M, Thurfjell L, Vandenberghe R, Ourselin S, *et al*. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. Eur J Nucl Med Mol Imaging 38(6): 1104-19 (2011).

[10] Samuraki M, Matsunari I, Chen WP, Yajima K, Yanase D, Fujikawa A, *et al*. Partial volume effect-corrected FDG PET and grey matter volume loss in patients with mild Alzheimer's disease. Eur J Nucl Med Mol Imaging 34(10): 1658-69 (2007).

[11] Forster S, Yousefi BH, Wester HJ, Klupp E, Rominger A, Forstl H, *et al*. Quantitative longitudinal interrelationships between brain metabolism and amyloid deposition during a 2-year follow-up in patients with early Alzheimer's disease. Eur J Nucl Med Mol Imaging 39(12): 1927-36 (2012).

[12] Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. Phys Med Biol 57(21): R119-59 (2012).

[13] Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. J Nucl Med 39(5): 904-11 (1998).

[14] Muller-Gartner HW, Links JM, Prince JL, Bryan RN, McVeigh E, Leal JP, *et al*. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. J Cereb Blood Flow Metab 2(4): 571-83 (1992).

[15] Leung KK, Barnes J, Modat M, Ridgway GR, Bartlett JW, Fox NC, *et al*. Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. NeuroImage 2011 55(3): 1091-108 (2011).

[16] Mikheev A, Nevsky G, Govindan S, Grossman R, Rusinek H. Fully automatic segmentation of the brain from T1-weighted MRI using Bridge Burner algorithm. J Magn Reson Imaging 2008 27(6): 1235-41 (2008).

[17] Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 17(3): 143-55 (2002).

[18] Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, *et al*. A hybrid approach to the skull stripping problem in MRI. NeuroImage 22(3): 1060-75 (2004).

[19] Gray KR, Wolz R, Heckemann RA, Aljabar P, Hammers A, Rueckert D, *et al*. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage 60(1): 221-9 (2012).

[20] Sanabria-Diaz G, Martinez-Montes E, Melie-Garcia L. Alzheimer's Disease Neuroimaging I. Glucose metabolism during resting state reveals abnormal brain networks organization in the Alzheimer's disease and mild cognitive impairment. PLoS One 8(7): e68860 (2013).

[21] Jack CR, Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, *et al*. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Mag Res Imaging 27(4): 685-91 (2008).

[22] Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, *et al*. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 30(2): 436-43 (2006).

[23] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17(1): 87-97 (1998).

[24] Ashburner J, Friston KJ. Unified segmentation. NeuroImage 26(3): 839-51 (2005).

[25] Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, *et al*. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp 19(4): 224-47 (2003).

[26] Rousset OG, Collins DL, Rahmim A, Wong DF. Design and implementation of an automated partial volume correction in PET: application to dopamine receptor quantification in the normal human striatum. J Nucl Med 49(7): 1097-106 (2008).

[27] Rousset OG, Deep P, Kuwabara H, Evans AC, Gjedde AH, Cumming P. Effect of partial volume correction on estimates of the influx and cerebral metabolism of 6-[(18)F]fluoro-L-dopa studied with PET in normal control and Parkinson's disease subjects. Synapse 37(2): 81-9 (2000).

[28] Watanabe H, Andersen F, Simonsen CZ, Evans SM, Gjedde A, Cumming P, *et al*. MR-based statistical atlas of the Gottingen minipig brain. Neuroimage 14(5): 1089-96 (2001).

[29] Gutierrez D, Montandon ML, Assal F, Allaoua M, Ratib O, Lovblad KO, *et al*. Anatomically guided voxel-based partial volume effect correction in brain PET: impact of MRI segmentation. Com Med Imaging Graph 36(8): 610-9 (2012).

[30] Brendel M, Hogenauer M, Delker A, Sauerbeck J, Bartenstein P, Seibyl J, *et al*. Improved Longitudinal [18F]-AV45 Amyloid PET by White Matter Reference and VOI-based Partial Volume Effect Correction. Neuroimage 108: 450-9 (2015).

[31] Frouin V, Comtat C, Reilhac A, Gregoire MC. Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness. J Nucl Med 43(12): 1715-26 (2002).

[32] Strul D, Bendriem B. Robustness of anatomically guided pixel-by-pixel algorithms for partial volume effect correction in positron emission tomography. J Cereb Blood Flow Metabol 19(5): 547-59 (1999).

[33] Quarantelli M, Berkouk K, Prinster A, Landeau B, Svarer C, Balkay L, *et al*. Integrated software for the analysis of brain PET/SPECT studies with partial-volume-effect correction. J Nucl Med 45(2): 192-201 (2004).

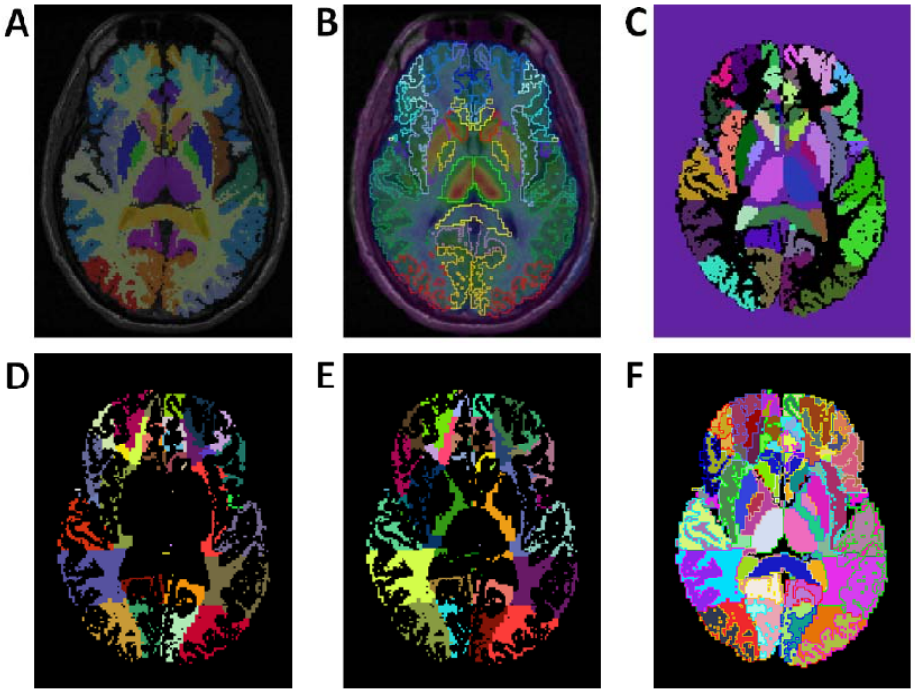
[34] Jack CR, Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, *et al*. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9(1): 119-28 (2010).

[35] Schmidt R, Ropele S, Pendl B, Ofner P, Enzinger C, Schmidt H, *et al*. Longitudinal multimodal imaging in mild to moderate Alzheimer disease: a pilot study with memantine. J Neurol Neurosurg Psychiatry 79(12): 1312-7 (2008).

Received: February 05, 2015 **Revised: October 10, 2015** **Accepted: October 16, 2015**

Supplement 1

Image processing / PVEC illustrated on axial slices



(A) T1w-MRI fused with grey matter segments. Segments are obtained by masking the atlas with either a desired grey matter probability threshold (0.3 in our case) or a binary mask deriving from segmentation. Previously acquired transformations from PET to MRI and MRI to atlas/MNI space are used inversely to enable analyses in native PET or MRI space.

(B) MR fused with PET and overlaying with outlined VOIs.

(C) In the first step of preparation for VOI-based PVEC processing, white matter VOIs are removed, as those regions are subdivided into smaller regions in the subsequent steps. Second, the remaining original grey matter VOIs (multicolored) are subtracted from the atlas mask. This operation therefore defines voxels that were removed from the atlas during grey matter thresholding, or were assigned to white matter regions. The resulting voxels are illustrated in black.

(D) The mask created from voxels defined in the previous step is intersected with the atlas. This step creates complementary regions from the atlas definition.

(E) Finally, complementary white matter regions are created by adding voxels which were previously in white matter regions, but do not have a specific definition at this stage of processing. These regions are added by outlining to the original VOIs as complementary VOIs.

(F) Final regions are passed to the PVEC algorithm.

Eidesstattliche Versicherung

Högenauer, Marcus
Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema:

Auswirkung von Referenzregionen, Partialvolumeneffekten und Segmentierungsartefakten auf die Quantifizierung der Positronen-Emissions-Tomographie bei der Alzheimer-Erkrankung

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Wolnzach, 5.11.2017 _____
Ort, Datum Unterschrift Doktorandin/Doktorand